International Journal of American Linguistics

VOLUME XXIII April 1957 Number 2

TESTING INTELLIGIBILITY AMONG YUMAN LANGUAGES

BRUCE BIGGS

AUCKLAND UNIVERSITY, NEW ZEALAND

0. Introduction
1. Compilation of tests
2. Administration of tests
3. Scoring and analysis of tests
4. Results
5. Conclusions

0. This paper records the results of a month's fieldwork among six Yuman tribes in Arizona and California during the summer of 1956. The project was suggested by C. F. Voegelin and made possible by the generosity of Indiana University, which supplied field funds, and the hospitality of the Museum of Northern Arizona whose Research Centre provided an ideal base.

In 1950 Voegelin and Harris[1] described a "test the informant" method of measuring the degree of intelligibility between related languages, a method which had been made practicable by the recent development of convenient recording equipment. Practical application of the method has been described for Iroquoian and Algonquian languages, in papers which also discuss certain difficulties involved in the testing and scoring of results.[2]

Briefly the technique involves asking an informant to respond in the investigator's language to a tape recording of a text in a language different from, but related to, the informant's own language. The response is scored against a master translation, and the result expressed as an index of the degree of intelligibility.

The closely interrelated Yuman languages are spoken by small tribes living in Northwest Arizona (Walapai, Yavapai, Havasupai), on the Colorado river north of the border (Mohave, Yuma), on the Gila river near Phoenix (Maricopa, Halcidhoma, Kavelcadom), on the Colorado delta (Cocopa, Halyikwamai, Kahwan) and north and south of the border in California (Diagueño, Kamia, Akwa'ala, Kiliwa). Kroeber describes the above linguistic divisions[3] mainly on the basis of shared lexical and phonological features, but also taking into consideration ethnographic data, and informant opinion. He further says that Halcidhoma and Kavelcadom were very close to Maricopa; that Maricopa is more than a dialect of Yuma having agreements with Mohave as against Yuma; that the Northwest Arizona languages form a close-knit linguistic group; that there are three dialects of Yavapai.

Because of the short time available the present study might have been confined to a consideration of the Northwest Arizona languages alone, but it soon appeared that the three languages concerned, Walapai, Yavapai and Havasupai were all very close. Accordingly the project was extended to include Yuma, Mohave, and Maricopa in the hope of introducing more variable data.

The aim of the study was to demonstrate by means of the "test the informant"

[1] C. F. Voegelin and Zellig S. Harris, Methods for Determining Intelligibility Among Dialects of Natural Languages, APS-P 95.322–29 (1951).

[2] Harold Hickerson, Glen D. Turner, and Nancy P. Hickerson, Testing Procedures for Estimating Transfer of Information Among Iroquois Dialects and Languages, IJAL 18.1–8 (1952), and Joe E. Pierce, Dialect Distance Testing in Algonquian, IJAL 18.203–210 (1951).

[3] A. L. Kroeber, Classification of the Yuman Languages, UCPL 1.21–35 (1948).

method the relative distances of the six languages mentioned, and to suggest the validity of the results by comparison with those arrived at by other methods.

1. I was able to compile a suitable Walapai test and a short Maricopa test before going into the field at all,[4] so that from the beginning collecting of texts and administering of tests was possible at one and the same session, which was desirable in view of the distances involved and the time available. Thus on the first field trip, which was to the Yavapai reservation at Prescott, it was possible to administer tests on Walapai and Maricopa, and also to elicit material to be arranged as Yavapai tests.

Tests were compiled from running texts recorded by the more fluent informants. By playing back phrase by phrase, and recording each phrase and its translation on a second machine, translations were obtained at the same session as the texts. Accounts of practical situations made the best test material that was obtained in this way, being translated more satisfactorily by the original informant, and consistently providing more directly comparable test results than say, moralising statements about Indian capacity for work, quarrelling, or liquor, all of which were obtained. To secure suitably oriented material informants were encouraged to give short autobiographical statements.

One folkloristic text was used in a test and no informant admitted to having heard the tale before; but in other circumstances this might have been unwise, if incidents in the tale were widely known, or if the language differed from the colloquial norm. Most informants were able to speak English though with varying degrees of fluency, and in only one instance was an elderly Yavapai woman unable to give an English translation of her own text. Informants who were unable to give spontaneous texts sometimes responded well in providing translations of a series of English sentences. It will be suggested later that such translations provide the best test material.

In arranging tests an endeavour was made to have two or three selections with differing lexical ranges for each language. So for Havasupai sub-test (a) a selection was made from a rather moralising account of a pow-wow,[5] sub-test (b) was autobiographical, and sub-test (c) consisted of sentences translated from English. The tests were transcribed in broad phonetic transcription and "pause to pause" sections were numbered for convenience in administering.

The nature of the project demanded that a considerable number of informants be used. About seventy Yuman speakers were approached at Flagstaff, Grand Canyon, Peach Springs, Clarkdale, Prescott, Parker and Laveen in Arizona, and at Needles and Winterhaven in California. Some were interested and anxious to be helpful, others were too occupied with their own affairs, or shy, or suspicious, and one was hostile. Thirty-two took some part in a test situation.[6]

2. From the first it was evident that to insist on rigid test procedures would mean restricting the experiment to the few informants who could be interviewed alone. Actually several of the tests were conducted with more than one informant present and my only restriction was against help from persons bilingual in the test language. On almost every such occasion one individual elected to respond in English while others present offered occasional help in Indian. The practical need for such an informal procedure with unsophisticated informants

[4] These were made available by Werner Winter. Both he and J. Hopkins also helped me considerably in contacting informants.

[5] This sub-test proved unsuitable because of translation difficulties and scores obtained on it were disregarded in the final analysis.

[6] My wife was very successful in overcoming the reticence that I found in marked contrast to the Polynesian friendliness with which I am more familiar.

is, I think, obvious; theoretically it can be justified by the same considerations that later determined the choice of the highest scores rather than the mean scores as indices of the amount of transfer (**3**). Anything, other than help from bi-linguals, that pushed the score towards the upper, presumably fixed limit of transfer, was regarded as desirable.

Every informant other than those who had supplied the test material and translations originally was first tested on his native language. The procedures and objectives of the study were explained before the tests were played, in an effort to arouse interest and to dispel the idea that an informant might be expected to have a large store of ethnographic information, a notion that had prompted several people to refuse to participate. Then the test was played phrase by phrase and the informant's response in English was written down. More than one playing of a phrase was permitted and the informant was allowed to alter his responses, the final version only being kept in the record. After a few phrases the informant was asked to identify the language as his own.

A similar procedure was followed with the other languages, which were not presented in a fixed order but rather languages were alternated which were considered to be close and distant from the informant's language. In particular it was considered necessary for the first non-native language presented to be one in which a large amount of transfer was expected. This encouraged informants to respond positively to each phrase and not be content with answering negatively when the content of a phrase was not immediately obvious. Incorrect responses were noted silently but correct translations were sometimes approved, especially when an informant was getting little information from a distant language. During each test the subject was asked if he could identify the test language. If he was able to do so correctly an effort was made to judge the extent of his acquaintance with it. If this was considerable the test results for that language were disregarded in the final compilation of results. It took about an hour to run through all the tests and it is considered that it would be difficult always to maintain interest over a longer period.

3. Scoring began after all testing was completed. The reliability of the method must depend to a considerable extent on the possibility of obtaining a "true" translation of each test as a basis for scoring. In this situation the problems posed by the investigator's lack of knowledge of the test languages and the informants' imperfect knowledge of English are added to the difficulties inherent in any translation. It was decided that such grammatical categories as person, tense-mode, and number would be ignored. While it was sometimes possible to ascertain when an informant was consciously substituting third person for first person in the text, it was not always so. In general it was noted that informants did not handle English tense-mode or number very well and rather than attempt to make allowance for various degrees of proficiency in handling English grammar the scoring was confined to words of substance.

All translations by native speakers of the language of a text were collated for the master translation and an effort was made to identify each English item in the original test. Then test results were scored according to the number of items corresponding to the master translation. Scores for each language were expressed as percentages of a perfect score represented by the number of items in the master translation. The numbers of scoring items in each language ranged from forty-one in the case of Mohave to eighty for Yuma. Twenty-nine indices of intelligibility between different languages and five between speakers of the same language were obtained in this way. (No tests were conducted for Yuma to Mohave,

TABLE I

Maricopa to Maricopa	96
Walapai to Walapai	96
—	
Yavapai to Yavapai	94
—	
Mohave to Mohave	84
—	
Havasupai to Havasupai	91
—	
Walapai to Havasupai	98
Havasupai to Walapai	91
—	
Yuma to Maricopa	98
Maricopa to Yuma	96
—	
Yavapai to Havasupai	83
Havasupai to Yavapai	78
—	
Yavapai to Walapai	96
Walapai to Yavapai	82
—	
Maricopa to Mohave	77
Mohave to Maricopa	67
—	
Mohave to Yuma	49
Yuma to Mohave	—
—	
Walapai to Maricopa	18
Maricopa to Walapai	14
—	
Mohave to Walapai	27
Walapai to Mohave	11
—	
Mohave to Yavapai	20
Yavapai to Mohave	13
—	
Yavapai to Maricopa	13
Maricopa to Yavapai	12
—	
Mohave to Havasupai	18
Havasupai to Mohave	16
—	
Yuma to Yavapai	15
Yavapai to Yuma	4
—	
Maricopa to Havasupai	11
Havasupai to Maricopa	10
—	
Yuma to Havasupai	9
Havasupai to Yuma	4

nor for Yuma to Yuma). The range of the scores obtained by speakers of a language on tests in the same language (see table I) suggest that a score of over 90 per cent between different languages indicates transfer equal to that between speakers of the same language.

In the final analysis of results the highest scores obtained on each test were taken as the measures of intelligibility. The assumption here is that the upper limit of understanding between any two languages is set by linguistic factors within these languages, and that such an upper limit will vary for different pairs of languages. If this is so then the upper limit of understanding is a suitable measure of language distance. This upper limit of transfer is most nearly approximated by the highest scores obtained on the tests. The downward range from the highest score, which may be considerable, is presumably due in great part to a number of other factors such as general intelligence, test motivation, facility in translating into English, and so on.

4. Following are the highest scores obtained on each test by a speaker of each language. Test languages are on the left, languages of the subjects being tested on the right.

In suggesting the positions of the six languages relative to each other it would be simpler to have a single index between two languages rather than the pairs of the above table. In this case it would not seem desirable to take the higher index of each pair because the assumption cannot be made that the optimum transfer from language A to language B is equal to the transfer from B to A. The above figures suggest that the degree of understanding of Yuma by Walapai, Yavapai, and Havasupai speakers is in each case greater than the reverse. This may be due to other than linguistic factors but there seems to be no *a priori* reason to expect transfer to be equal in both directions. Accordingly an average is taken of each of the pairs of indices in the above table. The results are

TABLE II

	Transfer per cent	Cognate per cent
Walapai–Havasupai	95	95
Walapai–Yavapai	89	91
Yavapai–Havasupai	80	92
Yuma–Maricopa	97	—
Mohave–Maricopa	72	85
Yuma–Mohave	49	—
Mohave–Walapai	19	63
Walapai–Maricopa	18.5	57
Mohave–Havasupai	17	63
Mohave–Yavapai	16.5	62
Yavapai–Maricopa	12.5	57
Yuma–Walapai	11	—
Maricopa–Havasupai	10.5	58
Yuma–Yavapai	9.5	—
Yuma–Havasupai	6.5	—

compared in table II with the relevant lexical cognate counts obtained by Winter.[7]

5. There is general agreement between the results arrived at by the two methods of "testing the informant" and "counting samenesses". Walapai, Havasupai, and Yavapai are placed together with some suggestion that the first two are closer to each other than either is to Yavapai. In southern Arizona Maricopa and Yuma are placed very close by the testing method though no results are available here for lexical samenesses. Both methods agree in placing Mohave between the other two groups but closer to the southern. Moreover these results agree with Kroeber's classification with the possible exception that Maricopa is closer to Yuma than he suggested. In this connection I had noted that one of my Yuma informants on hearing a Maricopa text said that it was a woman talking Yuma with a Maricopa accent. Later at Laveen a Maricopa hearing the same Yuma informant thought at first that he was speaking Maricopa.

There are difficulties in the use of the informant testing method where related speech communities are geographically contiguous but some of these difficulties, such as partial bilingualism for instance, are common to other methods for determining dialect distance. The informant testing method as used here however may also fail to distinguish closely related dialects because of translation difficulties faced by informants whose English is inadequate. Ideally the test material should be such that the best informants can score one hundred per cent when tested on their own language. Only then can scores of say, ninety per cent on another language be considered significant in marking differences. The sub-tests that were translated best and most consistently by native speakers of the test languages were unconnected sentences suggested by myself and translated from English by other informants. Such material also offered higher possibilities of transfer into other languages. It is suggested that confining the tests to such material would sharpen the technique, enabling close dialects to be distinguished more clearly.

At the lower extreme transfer between the northwest group and Maricopa was below 15 per cent in spite of sharing at least 50 per cent of cognates. It is likely that somewhere not too far below this figure transfer would tail off to nothing,[8] thus setting a lower limit to the usefulness of the method.

Within the limits of extreme closeness and a degree of distance yet to be determined the method seems to be at least as sensitive as the comparison of cognates alone. Its main advantages however are seen in the relative speed with which usable results can be obtained, and the possibilities of its use by investigators unacquainted with the languages being tested.

Throughout this paper I have used the

[7] Werner Winter, Yuman Languages I, IJAL 23.18–23 (1957).

[8] See Joe E. Pierce, Crow vs. Hidatsa in Dialect Distance and in Glottochronology, IJAL 20.134–36, where practically no transfer was obtained with texts in spite of over fifty per cent of basic cognates.

word 'language' to refer to any named communalect recognised as such by the speakers. Now it is possible to suggest that Walapai, Havasupai, and Yavapai are clearly dialects of a single language, as are Yuma and Maricopa. Mohave belongs with the southern group and might be regarded either as a dialect of Yuma-Maricopa or as a separate language. As the nearest Mohave settlement, at Parker, is geographically separated from the Yuma-Maricopa groups by more than a hundred miles it seems unlikely that there is shading off of differences between them. It becomes a question then of the point at which a culturally useful but perhaps linguistically arbitrary decision between dialect and language is made, and an answer might be determined by considerations beyond the scope of this paper.

VOICELESS VOWELS IN COMANCHE

ELLIOTT D. CANONGE
SUMMER INSTITUTE OF LINGUISTICS

0. Introduction
1. Contrast in identical environment
2. Contrast in analogous environment
3. Pre- and/or post-syllabic vocoids
4. Rigg's interpretation B

0. To date, several published articles on Comanche[1] have postulated phonemic voiceless vowels in syllabic position. In spite of these reports, however, the phonemic interpretation of such vocoids as voiceless vowels is still considered an improbability by some scholars. In order to allay the general doubt expressed by Jakobson, for example, in the first printing of his 'Preliminaries', as to the existence of phonemic voiceless vowels in any language, I supplied him with tapes containing such sounds. In his second printing, he added spectrograms[2] illustrative of the difference between Comanche /i/ and /I/ occurring on these tapes but in spite of the contrast therein, which can readily be seen by inspection, he retained his scepticism as to their phonemic interpretation: "Vowels are normally voiced. It is still questionable whether there are languages in which parallel to the consonantal opposition voiced vs. voiceless, there actually is a similar distinctive opposition of voiced and murmured vowels, as reported about a few American Indian languages, e.g. Comanche. Either the vocal murmur is not a distinctive feature and functions merely as a border mark, or it may be a concomitant of the tense-lax opposition."[3]

This article, therefore, is an attempt to supply more extensive data and argumentation on this point than previous authors have given, since to me it appears conclusive that, in Comanche, voiceless (or 'whispered') vowels in syllabic position are phonemic. The interpretation of phonetic voiceless vocoids in pre- and/or post-syllabic position will also be discussed.

1. The phonemic contrast of voiced and voiceless vocoids in syllabic position can be seen in their occurrence in identical environment. Here, surely, no conditioning factor is present. Note, for /e/ versus /E/, ʔu'kuʔ-okʷekʷaiʔu? *she went to render it*, (e.g. *lard*), ʔu'kuʔokʷEkʷaiʔu? *she rendered it and went on*. In these two examples, the stress, length, pitch, rhythm, and phonological junctures are the same. The grammatical structure is also comparable: each begins with the morpheme ʔu- *object, 3d per. sing.*, followed by the morpheme ku- *heat as an instrument*; following the stem in each is the morpheme -kʷa *movement away*, the morpheme -i *past tense*, and the morpheme -ʔu? *subject, 3d per. sing.* The stem ʔokʷe means *a flowing or liquification with purpose directed toward the motion indicated*

[1] Joseph B. Casagrande, Comanche Linguistic Acculturation: I, IJAL 20.140 (1954); Henry Osborn and William A. Smalley, Formulae for Comanche Stem and Word Formation, IJAL 15.93 (1949); Venda Riggs, Alternate Phonemic Analyses of Comanche, IJAL 15.229 (1949); William A. Smalley, Phonemic Rhythm in Comanche, IJAL 19.297 (1953).

My own analytical research on this particular feature in Comanche has continued intermittently since 1948, when the data were presented to a seminar of the Summer Institute of Linguistics. Numerous informants have been used, chief of whom is Mrs. Emily Riddles, Walters, Oklahoma, whose patience has been greatly appreciated. I am also indebted to Kenneth L. Pike for numerous leads and practical suggestions in the preparation of this paper.

[2] Roman Jakobson, Preliminaries to Speech Analysis, Acoustics Laboratory, MIT, Report no. 13, (1952), Second Printing, 52, Fig. 12.

[3] idem. 26.

in the following morpheme, whereas the stem ʔokʷE is a complex stem composed of ʔokʷe modified by the addition of the morpheme comprised phonetically of stem final unvoicing, and which has the semantic implication of *stem meaning resulting in action indicated by the following morpheme.*

Brought about by the occurrence of this same stem-final-unvoicing morpheme, and preceding -ki *movement toward*, similar contrasts between voiced and voiceless vowels can be observed, e.g. 'norïʙakiki?u? *he came to pack* versus 'norïʙakIki?u? *he packed and came on.* Again, no phonological or morphological conditioning factors are present. Similar contrasts can be seen in 'mayakeki?u? *he came to play* (e.g. *the piano*), 'mayakEki?u? *he played and came on*; ?u'hanikïki?u? *he came to fix it for him* versus ?u'hanikÏki?u? *he fixed it for him and came on*; ?u'ni?aciki?u? *he came to advise him* compared with ?u'ni?acIki?u? *he advised him and came on.* Another contrast preceding -kʷa can be observed in 'namaka?-mukikʷai?u? *he went to dress himself* versus 'namaka'mukIkʷai?u? *he dressed himself and went on.* -ki and -kʷa are interchangeable in any of the illustrations listed above.

2. Utilization of analogous data should prove profitable, contrasting /u/ /U/, /o/ /O/, /e/ /E/. Thus ?u'?uhtuki?u? *he came to give to her* contrasts with ?u'?-utUki?u? *he gave to her and came on*; 'nohkoki-?u? *she came to bake biscuits* with 'nokOki?u? *she baked biscuits and came on*; and ?u'-manahkeki?u? *he came to measure it* with ?u'manakEki?u? *he measured it and came on.* These same stems occur with similar resultant contrasts preceding -kʷa.

Nothing in these word pairs is of such a nature or in such a position as to make a non-phonemic conditioning variation from voiced to voiceless a characteristic of the vowel. The theoretical possibility of /U/ /O/ being conditioned by voiced equivalents occurring in contiguous syllables preceding these voiceless vowels is partially eliminated by the occurrence of /E/ in a similar construction in which the preceding voiced vowel is not identical with it, that is, /a/. Instances of vowel assimilation have been observed, e.g. -ru?I *future tense* occurs as -ro?I in 'nohkoro?I?u? *she is going to bake biscuits.* Here, however, a change of vocalic timbre is involved rather than a change of voiced to voiceless. A more pertinent analogy may be noted in 'sikItï?u? *it is here*, 'sokOrï?-u? *it is over there*, 'sukUrï?u? *it is out of sight* and might seem to imply some sort of vowel agreement. However, e, indicating *scattered items*, occurring in a similar construction, results in 'sekItï?urïï *they are here and there*, not *'sekEtï?urïï.

Furthermore, limited instances of the occurrence of /U/ and /O/ in environments in which neighboring voiced vowels are not of the same vocoidal timbre with them have been observed, e.g. ?u'hotUkonïï *its hindquarters* and 'tuukOnenï?'ʙuni? *I saw a black cradle*, in which the voiceless phoneme /U/ is preceded and followed noncontiguously by voiced /o/, and in which the voiceless phoneme /O/ is preceded and followed noncontiguously by voiced /u/ and /e/ respectively.

The influence of neighboring voiced vowels as a criterion for predicting the distribution of the particular vocalic timbre of the voiceless vowels can be ruled out since voiceless vowels occur in contrast in identical environment where no conditioning element is present, as in '?ica?nï'rïïpI *this is my stone* compared with '?ica?nï'rïïpE *this is my mouth*; and in nonconditioning analogous environment, e.g. '?ica?nï'kïïkA *this is my onion*, '?ica?nï'hïïkI *this is my arbor*, '?ica?nï'?eekO *this is my tongue.*

In the analogy '?uhtu-, '?utU-; 'nohko-, 'nokO-; the presence or absence of /h/ (phonetically [U] or [O] in these instances) as a possible conditioning factor can be discounted since /h/ does occur (though rarely) contiguous to a syllable containing a voiceless vowel syllabic, as in nï'?uhtU *give it to me* compared with 'kimA *come*! or '?ica?nï'-

wahcĬ *this is my rib* versus '?ica?nï'nakI *this is my ear.*

It may be noted here, furthermore, that there is frequently morpho-phonemic alternation between a pre-consonantal /h/ and a post-consonantal voiceless vowel preceding certain suffixes. Compare the stem alternants 'pohpi- and 'popI- before the suffix -tu?I *future tense* and -ci *dependent completed action by the subject* in the illustrations 'pohpitu?Inï? *I am going to jump* and 'popIcinï?'miaru?I *having jumped, I will go.*

The contrast /a/, /A/ is extremely rare due to the low frequency of occurrence of /A/ and its frequent morpho-phonemic alternation with /Ĭ/. Whereas the analogies 'pohpi-, 'popI-; 'co?me-, 'co?mE-; '?uhtu-, '?utU- preceding -ki, -k^{w}a and other morphemes are common, the analogy 'tĭhka-, *'tĭkA- would be erroneous. Rather, the second form would be 'tĭkĬ-. However, preceding -ci *dependent completed action by the subject* and ci, a syllable in the stem 'ni?aci- *to advise*, a valid contrast can be observed, e.g. 'mi?Acinï?'tĭhkaru?I *having gone, I will eat* versus ?u'ni?aciki?u? *he came to advise him.*

The preceding illustrations have given crucial contrastive data for each of the six vowel pairs. The following miscellaneous sets of forms supplement those given above and the phonemicization postulated in reference to them.

In the first set of examples, the contrast between /i/, /I/ occurs following stops and preceding the identical morpheme -nïï *plural*, e.g. 'papinïï *heads*, 'nakInïï *ears*, 'tĭʙoowapInïï *teachers*, 'to?Inïï *pipes*. Note that it was a pair of this type which Jakobson used for his spectrographic analysis: 'tĭrï?aiwapInïï *hired hands* and 'tĭpinïï *stones*. At the time the tape was sent to him, the minimal pairs quoted above had not been discovered and none occur in the data of Riggs and Smalley. Nevertheless, the earlier data of contrast in analogous (non-conditioning) environment appeared to all of us conclusive even before the newer minimal pairs had been discovered and which have been included above.

In a specialized verb compound composed of noun plus verb, /i/, /I/ contrast preceding juncture in the set 'papiwïnĬkatïnï? *I have a headache* compared with 'nakIwïnĬkatïnï? *I have an earache.*

As the first syllable of a two syllable stem, the contrast /i/, /I/ may occur in stem syllables *not* contiguous to another morpheme following, as in the set 'tĭk^{w}iʙukitï?u? *he is whipping* (*something*) compared with ?u'kĭk^{w}Iya?etï?u? *he habitually tastes it.* Here, the stem in the second instance is -k^{w}Iya- so that the unvoicing is not a product in this instance of a change at morpheme boundaries. Possible conditioning by the morpheme kï- *use of mouth or teeth* preceding, can be discounted by 'kï?ok^{w}etï?-u? *he is slobbering*, in which the vowel in the syllable following kï- is voiced, i.e. /o/.

Though the morphemes -kï *benefactive* and -ki *movement toward* differ morphologically, in the following set they present a phonologically nonconditioning environment in which to contrast /e/, /E/, e.g. ?u'-wïhtok^{w}ekïtï?u? *he is killing for her* versus ?urii'wïhtok^{w}Ekitï?u? *he is killing them and coming on.*

The next sets of illustrations show contrast between /ï/ and /Ĭ/. They can be observed, for example, preceding -ka *stative* and -?e *habitual* in ?u'karïkï?etï?u? *she habitually cooks it* compared with ?u'karïkĬ-katï?u? *she is cooking it.* Note also the occurrence of voiced /ï/ preceding -ru?I *future tense* compared with the voiceless phoneme /Ĭ/ within a complex verb stem, as in the constructions 'pitïru?I?u? *he will arrive* and 'patĬrïyaai?u? *he drowned.* Preceding the dependent clause clitic -ku *dependent action being performed by secondary subject* and -ka, a like clitic denoting *completed dependent action*, a similar contrast is noted, e.g. '?uhka?u'karïkïkuma?'mi?a?I *as she was boiling that, he went* versus '?uhka?u'karïkĬkama?'mi?a?I *when he boiled that, he went.*

Contrast in a slightly different type of situation can be seen in the following instances in which /ĩ/, /Ĭ/ occur in syllables immediately preceded by morphemes, e.g. following nï- *person* and wï- *instrumental* in 'nïpïkarĩ?u? *he is burying someone* and ?u'wïpĬkarĩ?u? *he is cutting it open*. However, stem syllables do not always unvoice following wï- as in ?u'wïnucai?u? *he broke it up* compared with ?u'wïnĬkĩ?i?u? *he tightened it*. The phonetic similarity of the syllable ka immediately following the contrast in the first set presents a valid nonconditioning environment in which to compare the occurrence of /ĩ/, /Ĭ/. Furthermore, ka in each instance is a part of different unrelated stems (-pïka- *to bury*, -pĬka- *to cut open*) so that no morpheme boundary is involved.

Another type of analogical data is herewith mentioned, e.g. ?u'wïminaru?I?u? *he will break it* compared with ?u'wïmInaru?-I?u? *he will get sick over it*. The second example is written phonetically as [?u'wïMInaru?I?u?]. This constitutes a legitimate contrast since the allophones [m], [M] are conditioned by their relationship to vowels and not vice versa. That is to say, the total distribution of the phones [M], [N], [R], [W], [Y], [ʀ], [k^w], as opposed to their allophonic voiced equivalents, is preceding voiceless vowels. These voiceless consonant phones do not contrast with their voiced equivalents at any point, whereas voiceless vowel phonemes do contrast with their voiced vowel phoneme equivalents following /p/, /t/, /k/, etc. Other examples are ?u'haniki?u? *he came to do it* compared with ?u'hanIki?u? *he did it and came on*; ?u'-rïrïnitĩ?u? *he told him to do something* versus ?u'rïrĬni?i?u? *he planted it*; and 'norïʙaki?u? *she packed up* compared with 'norĬnaki?u? *she came to make a bed*.

3. Phonetic voiceless vocoids in nonsyllabic position are phonemically interpreted as consonantal /h/ because of their noncontrastive phonetic characteristics. In postposed position, e.g. ['paAkipĬ], /'pahkipĬ/ *leather*, ['wiInu], /'wihnu/ *then*, ['noOkopĬ], /'nohkopĬ/ *biscuits*, ['paAMU], /'pahmU/ *tobacco*, ['seEka], /'sehka/ *those here and there*, ['tïĬkapĬ], /'tïhkapĬ/ *meat*, such vocoids always occur contiguous to, and have the same vocalic timbre as the voiced vowel syllabic. Such sequences as *aĬ, *iU, *oE, with diverse timbre between voiced and voiceless, are structurally impossible in Comanche. In [cvV] syllables, the voiced vowel is clearly heard as the peak of sonority. Thus, the distribution of /h/ in 'pohpi- *to jump* parallels the distribution of the consonant /?/ in 'na?nïmïnü *relatives*.

Similarly, pre-syllabic voiceless vocoids in utterance initial syllables, e.g. ['UuuUpI], /'huuhpI/ *tree*, ['Aa?nii], /'ha?nii/ *beaver*, ['Ĭïïkɪ], /'hïïkI/ *brush arbor*, more closely parallel consonant distribution since all consonants except /ʙ/ and /ʀ/ occur utterance initially, whereas voiced vowels do not. In addition, the vocoidal timbre of the pre-syllabic element here is completely conditioned by that of the following voiced vowel so that no vocoid sequence such as *Ĭa, *Ui, *Eo ever occurs.

Pre-syllabic voiceless vocoids also occur in the middle of words following /?/, e.g., as in the sequence cv?Vv in ['na?ĬïmuukU], /'na?hïmuukU/ *safety pin*. Interpretation of [V] as /h/ in this position reasonably follows since this position is otherwise normally occupied by consonants. For example, note 'cï?cïki?u? *he strangled*, ?u'ro?nikai?u? *he put it away*. Here again, the voiceless vocoid in [cv?Vv] is completely conditioned by the following voiced vowel.

Phonetic voiceless vocoids occurring between voiced vowels have no distinguishing vocalic timbre, being phonetically colored at beginning and ending by the respective contiguous voiced vowels, as in 'nohitu?Inĩ? *I am going to play*. Here, the phoneme /h/ is phonetically a vocoidal slur from a

rounded voiceless vocoid to an unrounded voiceless vocoid but with the total slur appearing to take no more time than another voiceless continuant consonant such as /s/. Thus, the illustration given here could be written phonetically in some such way as ['noOIi-], but without the implication of an additional mora of length. In this position also, there are no contrasting vocoid sequences such as *aOa, *aIa, *aUa.

Other characteristics have been observed which differentiate between phonemic voiceless vowels and phonetic voiceless vocoids interpreted as /h/. Voiceless vowel phonemes occur in open syllables only, as seen in 'kupIta? *a light*, 'pukukĬni *a stable*, 'naʙUsa? *a chain*, whereas /h/ occurs contiguous to a voiced syllabic, opening or closing the syllable. In general, voiceless vowels maintain a full mora of length, though there often seem to be perceptible differences of length in reference to the different voiceless vowels occurring. Thus, /Ĭ/ of 'puhakĬni *a church* appears to be shorter than /I/ of 'kupIta? *a light*. The mora value of /h/, on the other hand, may vary considerably, not in accordance with its vocoidal timbre, but under various conditions of stress, deliberation, speed of speech, style, etc. Thus, /h/ in 'wihnu *then* ranges from a sharp vocoid [I] in slow deliberate utterance to mere breathiness of almost no mora value in more rapid speech.

Voiceless vowels never occur in syllables preceded by pause or in stressed syllables. /h/ may occur in either, e.g. 'huuhpI *tree*.

4. Riggs[4] has stated a preference for her Analysis B in which consonant clusters resulting from post-syllabic h plus consonant are treated as unit phonemes, choosing to eliminate hc clusters at the price of nine new phonemes (or eleven phonemes, since hʙ and hy, not listed in her paper, also exist). To be consistent, one would also have to postulate a series of pre-glottalized consonants for ?s, ?m, ?y, etc., which would reduce the number of consonant clusters to zero but would produce thirteen more new phonemes. All possible combinations of ? followed by a consonant occur except ??.

It is not economical to introduce twenty-three new phonemes into a system in order to eliminate two well-balanced types of consonant clusters, e.g. ?c and hc. In both clusters, the first element is a glottal consonant, namely a stop and a fricative.

Such an analysis, furthermore, disregards the perceptible division between syllables which occurs even in rapid speech, and, in exaggerated slowdown, informants show no tendency to treat hc as a close-knit unit but, rather, introduce a sharp break between the h and the consonant following. It is by no means impossible to assume that a single phoneme can be ambi-syllabic. Numerous such instances may be found but, as between alternate analyses, that one is to be preferred which interprets phonemes as mono-syllabic, other things being equal.

The interpretation of hc as a single phoneme results in needless grammatical complication. For example, the utterance translated *just nothing* may occur with alternate nonsignificant word order, e.g. 'nah'kehenA or 'kehena'nah. Under Rigg's Analysis B, in which hk constitutes a single phoneme, the allomorph of the morpheme meaning *just* is 'nah (three segmental phonemes) in the second citation but in the first instance, the alternate is 'na plus a feature (pre-aspiration) of the following phoneme. If, on the other hand, the alternant is assumed to be 'na (without pre-aspiration) in the first example, then 'kehena *nothing* of the second utterance has an alternant ʰ'kehenA in the first. The total distribution of 'na and ʰ'kehenA is the strict limitation of their occurrence contiguous to each other. Considering hc as a consonant cluster, however, would eliminate all such allomorphic problems and greatly simplify the analysis of the grammatical structure.

[4] Riggs, op. cit.

THE NATURE AND USES OF LINGUISTIC TYPOLOGIES

JOSEPH H. GREENBERG
COLUMBIA UNIVERSITY

1. Preliminary remarks
2. The logic of typological classifications
3. Linguistic typologies
4. Use of typologies

1. As contrasted with the other two main methods of linguistic classification, the areal and the genetic, typological procedures have tended to an uncertain and marginal status in linguistic science. In view of recent stirrings which appear to indicate renewed interest in typological procedures in linguistics as well as other fields, it seems appropriate to consider the general logic of typological classifications with particular reference to linguistics, to survey the kinds of linguistic typologies both potential and actual, and to evaluate the possible uses of typological analyses. This latter consideration is of particular importance in view of the common imputation of arbitrariness to typological as opposed to genetic classifications in linguistics.

It is of some interest to note that the term 'type' in the period prior to modern theories of evolution had a connotation opposite to arbitrariness to the highest possible degree. A type was a class sharing essential, not merely accidental, properties and was a true constituent of the very ground plan of the universe. With the appearance of evolutionary interpretations, however, the explanation of similarities as the result of common origin led to a new conception of classes, whether of species in biology or languages in linguistics, as groups whose members were connected through the historical process of generation. With these discoveries and the appropriation of the terms 'family,' 'relationship,' etc., in linguistics as the analogues of biological processes, only genetic classifications came to be considered as truly founded in the nature of things insofar as they reflected actual historical relationships. The word 'type,' then, became a kind of relict applied to all classifications involving non-historical criteria and which therefore were, by their very nature, felt to be arbitrary. The one prominent attempt to apply typological criteria during the nineteenth century, the well-known classification of languages into isolating, agglutinative and inflective in the attempt to avoid the charge of arbitrariness was either given an untenable evolutionary interpretation or fell into a basic confusion of typological and genetic considerations by the assertion, for example, that all agglutinative languages might be expected to be related.

In fields other than linguistics the meaning of the terms 'type' and 'typology' have been roughly similar to those just outlined for linguistics, and analogous problems arise regarding the formulation of non-arbitrary criteria for classification and the possible significance of typological resemblances. Such terms as 'feudal society,' 'patrilineal land organization,' 'industrial society,' etc., refer to types of societies which may well occur in historically independent instances, and questions may be raised concerning the precision of the criteria utilized and the usefulness of comparing, for example, feudal China with feudal Europe. In similar fashion, archeologists have traditionally used designations such as Mousterian for implements of certain forms wherever they appear and regardless of chronology and presence or absence of concrete historical connections. Some discussions have emphasized the search for typological criteria which are non-arbitrary in the sense that they possess real historical

implications, but it would appear that the prevailing use of the term in practice continues to be close, in principle, to its employment in linguistics.[1]

2. Any kind of classification involves the consideration of two factors: a set of entities which are to count as the individuals in the universe subject to classification, and one or more attributes in accordance with which each individual can be placed in a class or assigned a rank order or an absolute number as a measure of the extent to which the given individual possesses the attribute.

What is an individual for purposes of a particular classification must be defined in each instance. There are no absolute individuals, and what functions as an individual in one classification may function as a constituent of an individual or as a class of individuals in another classification. For example, in race classification we may take human beings or whole populations defined, perhaps, by geographical boundaries. In linguistics, the language is usually the individual entity figuring in classification typological or otherwise; but as will appear later, there are other possibilities. A first requirement for rigor in any classification is the definition of the entities to be considered as individuals. Language, of course, is a somewhat vague term; but major difficulties do not seem to arise from this fact.

Attributes may be classified as predicates, comparative concepts, and numerical concepts.[2] Predicates are attributes whose presence in a given individual is susceptible of a yes-no answer. Logically, it is of predicate form, whence the name, e.g., x is red. We often encounter a number of predicates referring to the same attribute; for example, skin color in the early Blumenbach classification of races into black, brown, red, yellow, and white; but these are always logically restatable as choices between black and not-black, brown and not-brown, etc. It is this restatability which provides the logical basis for the contention that all linguistic predicates are binary.

A comparative concept is of the form x is —er than y. Since it contains two variables, it is, from the logical point of view, a relation. A comparative concept gives rise to a serial order rather than a segregation into mutually exclusive classes. An example of a comparative concept is hardness in minerals. X is harder than y holds if and only if x scratches y and y does not scratch x. A numerical concept assigns to each individual a number as an expression of the degree to which the individual is invested with attributes. A common example is weight as given in pounds. It is clear that, other things being equal, numerical concepts are more exact than comparative concepts and comparative concepts more than simple predicates. If it is convenient, a numerical concept can be reduced to a comparative and a comparative one to a simple predicate, though with loss of information in each case. For example, x weighs 12 pounds and y weighs 10 pounds can be reduced to x is heavier than y. If we call a person weighing more than 200 pounds heavy, then the numerical concept of weight can be reduced to a choice between the simple predicates heavy and not-heavy.

In linguistics, comparative concepts do not play a significant role in typologies. The distinction between predicates and numerical concepts as attributes can be illustrated by Sapir's and Greenberg's morphological typologies, respectively. In the latter scheme certain reductions of quantitative to predicate attributes were

[1] The attempt to give historical significance to typological concepts in archeology is especially marked in Alex Krieger, The Typological Concept, AA 9.271–87 (1944), but archeological terminological practice does not seem to coincide with Krieger's proposals.

[2] For a discussion of these three kinds of attributes see C. G. Hempel and P. Oppenheim, Der Typusbegriff im Lichte der Neuen Logik (1936) and P. F. Lazarfeld, Some Remarks on the Typological Procedures in Social Research, Zeitschrift fuer Sozialforschung 6.119–39 (1937).

suggested; for example, that the predicate polysynthetic be applied to a language if its index of synthesis is 3.00 or larger.

A further principle of classifying attributes leads directly to the distinction between typological and non-typological classifications. This principle pertains to the manner in which distinct individuals may be said to be the same or similar with reference to a particular attribute. This aspect of classification will be called mode. There are three modes: the formal, the functional, and the processual. In the widest sense, the distinction between the formal and functional modes rests logically on the difference between internal and external attributes. This distinction can be illustrated by the following example. Let us suppose that a factory produces two implements of cast iron which are indistinguishable, not as yet having acquired any differences resulting from use, e.g., scratches or stains. By 'indistinguishable' is meant that an observer would not be able to tell them apart consistently. Let us say that one belongs to a consignment to be sent to Cleveland and the other to Chicago. These two pieces of iron at this point share all internal predicates, but one has the external predicate 'to be shipped to Chicago' but does not have the external predicate 'to be shipped to Cleveland,' and the other piece of iron *vice versa*. Again, if one is being examined at time T by someone while the other is not, it will have the external predicate 'handled at time T.'

The classification of artifacts into cutting tools, scraping tools, etc., is based on external predicates involving relationships to other things and is therefore functional. Examples from biology would include a classification of species into fresh water, salt water, and land animals, and ecological classifications in general. In linguistics the predicates 'national,' 'substandard,' 'spoken by wavy-haired people' are all predicates in the functional mode. The term 'pidgin' as generally used is ambiguous, being applied partly because of internal properties, e.g., simplicity of morphological structure, and partly because of external properties, e.g., not being the first language of any population, used as an auxiliary means of communication. There is often, of course, some degree of causal connection between internal and external properties, which remains to be investigated, but the distinction between formal and functional classifications is, in general, a clear one. There may be cases in which what is first interpreted as an external predicate is later, with deeper knowledge, considered to be internal. For example, the fact that a person becomes involved in an automobile accident was generally regarded as an external property, but recent investigations tend to show that 'accident-prone' may be considered an internal property of certain persons.

The distinction between the formal and the processual modes involves two distinct methods of comparison, the difference between which is crucial in linguistics and certain other fields. In formal comparison a particular attribute is defined, and the individuals are assigned to the class depending on whether they have the attribute or not. It is thus a very straightforward procedure. If we compare processually, however, we do not demand formal coincidence in the properties used as criteria, but merely similarity. Such judgments of similarity are based on the possibility of the two individuals being plausible resultants of formally identical originals through differential changes. Therefore, although synchronic in appearance, such comparison always involves theories concerning possible kinds of changes—that is, processes—and is implicitly diachronic. We may define typological, historical, and functional classification as involving formal, processual, and functional modes of comparison, respectively.

A particular attribute applied typologically may be called a typological criterion. By a typology, however, we usually indicate a 'system' of distinct typological criteria,

each of which involves a separate attribute in an attribute space. Thus in various phonologic typologies, point of articulation and manner of articulation of consonants comprise two of the attributes in an attribute space in terms of which any consonant can be located. Traditional terminologies often imply elaborate, hitherto unanalyzed attribute spaces, usually vague and at points inconsistent in formulation. The process of discovering and reformulating in a more precise form the underlying attribute space of such an implicit typology has been called substruction.[3] In this sense the phonologic typologies of Trubetskoy, Hockett, and Voegelin are substructions of the rather vague terminology of traditional phonetics. In such a process another logical stage is sometimes reached in which the presence of certain relationships among the attributes themselves or classes of these attributes are employed as criteria. An example of the former is Trubetskoy's use of the term quadrangular and triangular for vowel systems and of the latter, some terms of Jakobson, Fant, and Halle's analysis; for example, grave and acute, each of which contains several classes which are distinct from the articulatory point of view.

Finally, we may distinguish language typologies in which the individuals of the universe to be typologized are languages, from linguistic typologies in general, of which they are a subclass. Ordinarily we are concerned with language typologies and other linguistic typologies; some of those which will be illustrated below would, perhaps, not be considered typologies in prevailing usage.

3. Starting, then, with language typologies, one principle of classification will refer to the above-mentioned division of attributes into predicates, comparative concepts, and numerical variables. Since comparative concepts in the case of linguistics hardly play a role, our fundamental distinction here will be between non-quantitative attributes (predicates) and quantitative attributes (numerical variables). Quantitative attributes, in turn, can be treated either on the basis of occurrences in the system (systemic) or in actual usage; e.g., texts (pragmatic). This is a general extension of the notion of lexicon vs. text frequency employed in studies of vocabulary frequency. Thus, to say that vowel-consonant ratio in Italian is 7:22 because there are seven vowels and twenty-two consonants in the phonologic system is a systemic quantitative statement, whereas to give the ratio of vowels to consonants in a text is a pragmatic quantitative statement. All typological criteria may be considered, then, to be non-quantitative, systemic quantitative, or pragmatic quantitative.

It is clear, also, that an important basis of classification has to do with the particular aspect of language with which we are concerned; for example, phonology or morphology. In fact, all typologies hitherto proposed have dealt only with a single aspect. It will prove convenient to distinguish six classes of typologies from this point of view: phonologic, morphologic, syntactic, those pertaining to canonic form, semantic, and symbolic.

The area of phonology, along with morphology, has been the most frequently treated from a typological point of view. The most common approach stems from Trubetskoy and is based on the presence or absence of certain phonetic features as principles of phonemic distinction in particular languages. Thus a comprehensive system such as that of Jakobson, Fant and Halle would assign every language in the world to one of a maximum of 2^n typological classes, where n is the number of principles of binary opposition enumerated. Actually, fewer than 2^n classes occur because the use

[3] This is the term employed by Lazarfeld (1937) and following him, A. Barton, The Concept of Property Space in the Language of Social Research (Glencoe) 1955.

of certain binary oppositions excludes that of certain others.

Non-quantitative typologies based on phoneme sequences are likewise possible and are implied by the common use of such terms as open syllable, closed syllable, consonant cluster, etc. Thus languages can be classified according to the syllable types or consonantal and vocalic sequences which occur. One simple scheme would divide languages into those which always have an initial non-peak consisting of one phoneme only, and those which allowed clusters in this position, and as the second axis of classification those which permitted phonemes after the peak and those which do not. Yokuts and Classical Arabic would then belong to the same class since in both languages every syllable contains a single non-peak and a maximum of one phoneme is allowed after the peak within the syllable.

An example of a systemic quantitative phonologic typology is the measurement of degree of symmetry of a sound system by calculating the number of combinations of sound features utilized to the total combinational possibilities of the phonemic system, as suggested by Saporta.[4] Here belong also measures of vowel-consonant ratio based on occurrence in the system.

The foregoing systemic typologies have pragmatic counterparts based on text counts. The consonant-vowel ratio is a simple instance of such a typological criterion. It is also possible to count the relative frequency of different syllable types or of phoneme sequences based on comparable features in different languages. After compiling frequency data for phonemes or phoneme sequences, it is possible to measure the entropy of the system and thus compare various languages from the point of view of entropy or economy in the use of the system of phonemes and phoneme sequences.[5]

Morphologic typologies, the next to be considered, have employed criteria based on the internal structure of the word. Thus the traditional nineteenth century typology considered such factors as the number of morphemes in the word and the existence or absence of irregular morphophonemic alterations, though without explicit statement. Sapir's discussion is essentially an attempt to define and reformulate in a more precise manner the implicit attribute space. Sapir's use of phrases such as slightly polysynthetic and strongly fusional comes nearest to a comparative concept found anywhere among linguistic typologies and, in so doing, replaces the nineteenth century predicate attributes by which languages were, for example, either analytic or synthetic but not more or less so. My typology is a pragmatic quantitative approach using Sapir's attribute space with some modification. A systemic quantitative approach is also possible for some of these indices, for example, the agglutinative; a comparison of the results of such an approach with corresponding pragmatic indices would be of interest in showing the manner in which the morphological resources of a language are employed in actual practice.[6]

Some of Sapir's categories and the corresponding indices in Greenberg's treatment skirt the borderline of syntax, e.g., the indices of concord and gross inflection. These are several possible types of a purely syntactic nature. One is the measurement of the degree of freedom of word order. One might take each word and consider it free or bound, depending upon whether it could be moved to another part of the sentence, other things being held constant, without changing the construction. The relative order of modifier and modified can

[4] Sol Saporta in Osgood and Sebeok, Psycholinguistics (1954).

[5] An example is written English in C. E. Shannon and W. Weaver, The Mathematical Theory of Communication (1949).

[6] This problem is raised in regard to Yokuts in Stanley Newman, Semantic Problems in Grammatical Systems and Lexemes: A Search for Method in Language in Culture, ed. H. Hoijer 82-91 (1954).

also be considered by both quantitative and non-quantitative means. This particular feature has been employed by Bally in his contrasting typological characterizations of French and German.[7]

A further area of research is possible in the domain of word classes. If objective cross-linguistically valid definitions could be developed, then languages might be classified by the presence or absence of particular classes. At present, essentially typological judgments of an imprecise nature are made in this regard. If, for example, there are languages without distinct noun and verb classes, which is subject to real doubt, then these languages would form a typological class against those that make the distinction. Similarly, there are said to be languages in which adjectives are absent as a class. Another type of study is that of relative word frequencies such as those conducted by Zipf. These, like some of the above-mentioned morphologic studies, lead to measures of entropy.

For want of an existing term, Hockett's phrase canonic form is here extended to all studies dealing with the phonemic make-up of grammatical units. The phonemic structure of morphemes, for which Hockett has employed this term, constitutes one subdivision of this general field. The triliteralism of Semitic root morphemes and the pattern of Penutian disyllabic roots with identical vowels in both syllables are examples of canonic form in Hockett's sense. Menzerath and his collaborators have developed this subdivision of typological studies.[8] Number of phonemes per morpheme or number of phonemes per word are possible subjects of investigation employing either pragmatic or systemic quantitative methods.

With more specific reference to the phonetic characteristics of the phonemes, we may have typological criteria by which languages are classified as having or not having vowel or consonantal harmony of particular types. An example of a systemic quantitative study here is my study of the patterning of root morphemes in Semitic and related languages.[9] Using the syllable as a phonologic unit, the traditional term monosyllabic language implies a quantitative ratio between numbers of syllables to words either pragmatically or systematically. These are but a few of the actual and possible types of study in the area of canonic form.

In semantic typologies we consider certain meaning categories and examine the varying methods by which languages express these meanings. An obvious instance is numerical systems. The classification of languages according to whether their terminology of numerals has a binary, quinary, decimal, duodecimal, or some other base is an example of a semantic typological procedure. Likewise, kinship systems may be analyzed by componential procedures as suggested by Lounsbury and then compared typologically.[10] Quantitative methods in this area are probably not feasible at the present time and perhaps would not accomplish anything not more easily attainable by purely qualitative methods. However, comparative concepts may here prove of some value. Thus one might want to be able to state that one language had a more elaborate and specialized vocabulary pertaining to zoological terms than another. Another area to be explored is comparative metaphor. For example, one might use a standard set of directional terms above, below, in back of, etc., and compare the extent to which languages utilize terms derived from body parts to express these ideas. Here a comparative scale, or even a systemic quantification, seems possible. Still another

[7] C. Bally, Linguistique générale et linguistique française 2nd ed. (Berne) 1944.

[8] P. Menzerath, Typology of Languages, Journal of the Acoustic Society of America 22.698–700 (1950); P. Menzerath and W. Meyer-Eppler, Sprachtypologische Untersuchungen I. (Lund) 1950.

[9] J. H. Greenberg, Patterning of Root Morphemes in Semitic, Word 6.162–84 (1950).

[10] F. G. Lounsbury, A Semantic Analysis of the Pawnee Kinship Usage, Lg 32.158–94 (1956).

subject of investigation is semantic vagueness and homonymy. Is it true that certain languages tolerate more homonymy than others? Vagueness might be measured for certain terms by tests in which a series of objects more and more unlike the average specimen of a class and tending towards a class designated by another term are presented to subjects to be named. The degree of intersubjective agreement and uncertainty can then be measured.[11] If one took a set of roughly equivalent terms for a number of different languages, a measure of average vagueness for each language could be calculated. There are, of course, many difficulties still to be overcome in investigations of this sort, but a beginning has already been made with color terms.[12] The writer's a priori expectation would be that no significant differences among languages would appear in this respect, but doubtless some genetically oriented psychologists would expect otherwise and would offer an evolutionary interpretation by which more 'primitive' languages would be expected to show greater areas of vagueness. A related problem is the extent of generic terms as opposed to specific terms in certain areas of lexicon for which no satisfactory procedure of measurement has as yet been suggested.

Thus far all criteria considered have involved sound without meaning or meaning without sound. Indeed it would be possible in practice to define typologies for linguistics on this basis since, typically, it is comparison of this kind which, by its disregard of one or the other factor, is, in general, without historical implications and provide a conveniently small set of typological classes. Thus, if languages in a phonological typology are called similar if they employ phonemes of tone, that is, are assigned to a class of tonal languages in disregard of the meanings of the forms in which the tones figure, then only two classes of languages result. If we specify the phonemes of tones as level or contour in addition only four classes result; in this case we have: 1. non-tonal, 2. level tones, 3. contour tones, 4. both level and contour tones. Even if we elaborate further, the number of classes will be small, and membership in the same typological class will have no necessary historical implications. Likewise, a purely semantic classification of languages into those which employ body part metaphors for spatial relations and those which do not, will involve meaning without sound since we will not consider the phonemic sequences by which the body-part terms are characterized. On the other hand, were we to classify languages into those which expressed the notion of 'dog' by the phoneme sequence 'dog,' those which expressed it by 'kane,' etc., we would arrive at thousands of classes and membership in the same class would have historical implications. A somewhat transitional zone is constituted by comparisons involving order and meaning. Order is formal principle but its possibilities of variations are few. Thus the fact that two languages both express pronominal subjects by a prefix to the verb has relatively insignificant historical implications and leads to a viable typology with a small number of classes. One might therefore define a typological criterion in language as one that does not employ sound and meaning simultaneously. However, apart from the fact that such a definition tends to obscure the points of contact between typology in linguistics and in other subject matters, it excludes at least one possible type of classification to be discussed shortly, the symbolic, which would be generally felt to be typological in nature. This possibility arises as follows: As we reduce the specificity of the requirements of phonetic resemblance of the form expressing some particular meaning, we diminish the historical significance of the resemblance and by the same token reduce the number of possible

[11] Max Black, Language and Philosophy (1949). See particularly pp. 52–3.

[12] E. H. Lenneberg and J. M. Roberts, The Language of Experience, IUPAL (1956).

types producing a more usable typological criterion. For example, we put all languages which have a nasal consonant in the word for 'mother' in the same class. The clustering of a large number of languages in the same typological class is a manifestation of what is usually called sound symbolism. A typology emerges when a sound symbolic connection has been determined in this manner for the equivalents of a number of concepts. A language may then be called more or less symbolic depending on the extent to which it belongs to the classes with the largest membership. All this represents a somewhat arduous procedure which has never yet been carried out in practice to this writer's knowledge. As with measures of semantic vagueness, certain psychologists would expect significant differences among languages in this regard, the more 'primitive' languages showing a higher degree of symbolism.

All the possibilities considered thus far have been language typologies as defined above; namely, typologies in which the individuals to be classified are languages. There is at least one possible typological area, however, in which diachronic changes rather than languages might figure as individuals. One talks of types of linguistic change, particularly phonetic change, and the usual terminology of conditioned change, unconditioned change, assimilatory change, dissimilatory change, etc., implies a typology. A class of diachronic changes based on conformity to the same typological criteria is usually called a process. Such typologies might be qualified also as diachronic as opposed to all those previously mentioned which are synchronic. Here again, although the individuals themselves are historical changes, membership in the same typological class involves no necessary historical connection among the changes themselves. For example, the Chwana languages in South Africa have undergone a series of consonantal shifts remarkably parallel to those described by Grimm's law for Germanic but obviously in a historically independent manner. Diachronic typologies are subject to much the same divisions as synchronic typologies: phonological, morphological, syntactic, semantic, canonic form, but probably not symbolic which, as has been seen, stands apart from the other categories in that it involves sound and meaning simultaneously.

4. The foregoing review is, of course, far from exhaustive. Moreover, it is realized that many of the types of studies discussed or suggested would not usually be called typological. Insofar as all of them involve comparison in which historical significance is entirely lacking or indifferent to the purpose of the study they exhibit a certain unity and the term typological seems the most suited by traditional usage to designate them in an overall way. For the most part, also, they involve individual typological criteria rather than a comprehensive typology defined by a significant attribute space. The application of individual criteria can be looked upon as a preliminary step to the development of full-fledged typologies in such instances.

The establishment of typological criteria sufficiently rigorous for application has a certain incidental usefulness, which is nevertheless of some importance for linguistics. In effect every descriptive scheme is typological and the very notion that the techniques of descriptive linguistics can be taught and can be employed on new and hitherto unknown languages requires the use of criteria, non-historic in their nature, which can be applied to any language. This is a commonplace in phonology, where terms such as stop consonant, vowel, etc. are employed cross-linguistically and automatically define typological criteria. In other areas of description, however, this is only partially the case. There is perhaps even the feeling that parts of speech and such units as the morpheme and the word involve different procedures for each lan-

guage and are therefore not comparable. The application of typological criteria and the ultimate development of rigorous typologies in these areas has therefore the heuristic value of stimulating us to seek for cross-linguistically valid definitions and procedures.

It should be noted that this does not involve imposing on any language a set of descriptive categories which do not conform to the facts of the language. We distinguish between the procedure and the results of the application of the procedure in individual instances. Thus the fact that there are universally applicable definitions of what is a vowel and for defining phonemes does not lead to the analytic result that all languages have the same phonemic vowel system nor would it prevent us from saying that a particular language had no vowels if such a language were to be discovered. In the same way it is clear that terms like 'adjective', 'word', and 'phrase' are of universal significance but that we are here usually far less clear about what they have in common in diverse instances than is the rule for phonological concepts. But the delineation of such procedures is forced upon us in typological work if languages are to be validly compared in these respects.

The formulation and application of individual typological procedures has a further value in that non-random distribution of languages among typological classes or in respect to quantitative attributes leads to the discovery of facts of universal scope concerning language and such facts constitute linguistic data of general significance for psychology, logic, physiology and perhaps other fields. For example, a typological study of subject, object verb and word order would doubtless show a non-random membership of languages in reference to the six logically possible arrangements and constitute a fundamental datum on the psychology of human thinking. It has been suggested in tentative fashion that the apparent predominance of suffixing over prefixing can be understood in the light of certain results of learning theory in psychology.[13] The frequency of types of vowel phonemic systems with only one low vowel and the rarity of such systems for mid and high vowels has physiological correlates. The distribution of the index of synthesis based on text counts of morpheme-word ratios in a normal distribution about a mean of approximately 2.2 is again a fundamental though as yet unexplained datum regarding human speech behavior.[14] Again in diachronic typologies the prevalence of anticipatory regular change over lag phenomena can be plausibly connected with certain well-known results of learning theory in psychology. These are but a few examples of possible results which may accrue from the widespread application of individual typological criteria.

A further step, as has been already noted, is the simultaneous application of a set of individual criteria, a process which can be conveniently conceptualized in terms of a multidimensional attribute space. The attributes will presumably be chosen in that they represent closely related aspects of the same general class of phenomena which may reasonably be expected to reveal significant mutual connections. The number of typological classes is in general larger, of course, for typologies than for individual criteria and consists of the product of the number of classes for each individual criterion. Here again, non-chance distribution of memberships in the various classes or, in quantitative typologies, significant positive or negative correlations among indices constitute significant data which require explanation. Unlike those mentioned above in the case of individual criteria, non-random

[13] C. E. Osgood and T. A. Sebeok, eds. Psycholinguistics, IUPAL (1954).

[14] J. H. Greenberg, A Quantitative Approach to the Morphological Typology of Language in Method and Perspective in Anthropology, ed., R. Spencer (Minneapolis) 1954.

distribution of languages among typological classes suggests in the first instance a connection among linguistic traits themselves, but such relationships may ultimately require explanation by reference to non-linguistic psychological, logical or other factors. To illustrate, the presumed lesser frequency of prefixing as opposed to suffixing was mentioned above. It has been further suggested not only that prefixing is less frequent but that it is more irregular than suffixing. This would manifest itself in a multidimensional typology by the rarity of prefixing-agglutinating languages as compared with prefixing-fusional. Likewise, Schmidt's contention that there are languages which in general put the modifier before the modified and others which place the modified before the modifier is in effect a multiple series of suggestions of the form: If in a given language the adjective is placed before the noun, it is likely with more than chance probability that adverbs are placed before the verb, etc.

In all these cases, then, we have a clustering of certain combinations of attributes while for others few or no languages exist which exhibit the combinations in question. In such instances a few salient types may emerge by a process discussed by Lazarsfeld under the name of reduction.[15] This can be accomplished by merging similar classes and disregarding classes with extremely small membership.

Finally, typologies can be utilized as a tool in the study of changes of types. The nineteenth century was generally connected in the minds of its adherents with a universal evolutionary scheme of change of type from isolating to agglutinative to inflective. There was even an attempt by Max Mueller at one time to connect these three types with hunting, pastoral and agricultural stages of economy, respectively, an assertion not repeated in later writings or accepted by any other writer. The question here raised does not concern such unilineal schemes but rather a careful study of changes of type in order to establish what limitations in the successions of types exist. Such knowledge clearly adds to our understanding of linguistic historical change and our predictive power since from a given synchronic system certain developments will be highly likely, others have less probability, and still others may be practically excluded.

The obvious method to be employed is the typological comparison of the same language at different chronological periods and the comparison of genetically related languages. The first results of such comparison in the latter case is a test of the phenomena of drift, or similar changes in related languages. Where the indices are quantitative a measure of this drift becomes possible. For example, in my morphological typology, for all of the ten indices changes in the same direction occur for Sanskrit-Persian, and Anglo Saxon-English. Additional unpublished data made available by Samuel Martin show the same results for Latin-French. The comparison of historically independent cases in different parts of the world would show whether there are overall tendencies in such matters or whether this is an Indo-European peculiarity.

[15] P. F. Lazarsfeld, Some Remarks on the Typological Procedures in Social Research, Zeitschrift fuer Sozialforschung 6.119-39 (1937).

SIX STATEMENTS FOR A PHONEMIC INVENTORY

C. F. VOEGELIN
INDIANA UNIVERSITY

Preface

I. Consonant types and linear consonants (LC)
II. Consonant subtypes and additive components (AC)
III. Vowel types and linear vowels (LV)
IV. Vowel subtypes and additive components (AC)
V. Ratios of LC to LV
VI. Ratios of consonant AC to vowel AC

A published sample of almost a score of languages is now typologized in the same way for consonant types and vowel types and some ratios between the two; namely, Maori (Malayo-Polynesian), Hidatsa and Winnebago (Siouan), Shawnee (Algonquian), Choctaw (Muskogean), Havasupai (Yuman), Navaho B (Athapaskan), Chontal (Mayan), Tarascan (unclassified);[1] and also Waiwai (Cariban), Chatino (Zapotecan), Totonac (Macro-Penutian), Witoto (Macro-Tupi-Guarani), Aguacatec (Mayan), Amuesha (Arawakan), Yagua (unclassified), Chipewyan and Navaho A (Athapaskan), Tlingit (Na-dene).[2] Before typologizing it is necessary to phonemicize, and if two phonemicizations of the same language differ, as do Navaho A (essentially Sapir and Hoijer's phonemicization) and Navaho B (Ken Hale's phonemicization), then the resulting typologies also differ.

In general, the prerequisite phonemicizations aim at giving a maximum description within a unique system framework. They certainly give more than do the derivative typologies; for example, they give the distribution of unit phonemes in terms of the allophones constituting the unit phonemes.

In general, the derivative typologies aim at giving a minimum description—which amounts to nothing more than a phonemic inventory, if we disregard the arrangement of the inventory—within a framework devised especially for comparability. Once the phonemic system of a given language is arranged in terms of the framework, it may be readily compared with every other language in the world similarly typologized. Comparability data of this order are only supplementary to the isomorphisms obtainable in comparative linguistics; in contrast, they may serve as primary data for work in areal linguistics, though of course they need not be restricted either to one language family or to a group of related and unrelated neighboring languages (areal linguistics).

The languages in the sample mentioned above are similarly typologized, each in terms of six statements, as follows.

I. The first statement is concerned with linear consonants (LC). For languages which include only linear consonants, all consonantal unit phonemes (C) are linear consonants (LC). We exemplify, to begin with, only from such languages in which the sum of LC = all C. We give the following information for each language exemplified: (a) the sum total of all LC; (b) the occurrence of these LC in numbered articulatory areas, from the lips (1) to the larynx (6); (c) from information given under (b), it will be possible to compute a type formula of proportional matching between stops (S), fricatives (F), nasals (N), liquids (L) or nasals and liquids (NL) at one or more

[1] John Yegerlehner, Florence M. Voegelin, and others, Frequencies and Inventories from Nine Languages, IJAL 23.85-93 (1957).

[2] C. F. Voegelin, Linear Phonemes and Additive Components, Word 12-3.00000 (1957).

areas specified by numbers. Our typological notation by numbers (1, 2, 3, 4, 5, 6) and by letters (S, F, N, L) frees us from the need of making traditional charts of phonemes in rows and lines.

(a) Maori shows 9 linear consonants (LC), Hidatsa 10 LC, Shawnee 11 LC, Waiwai 12 LC, Totonac 13 LC, and Havasupai 15 LC.

(b) The articulatory areas are numbered 1 for labials; 2 for dentals and alveolars (excluding affricates with [-s] and [-š] off-glide); 5 for velar consonants; and 6 for laryngeal consonants, including all varieties of h, and the glottal stop (but not glottalized stops). Though areas 1, 2, 5 and 6 are labelled in terms of articulatory 'position', differences between a stop at 2 (= /t/) and a stop at 5 (= /k/) also reflect 'kind' of articulation, since /t/ is not only more fronted than /k/ ('position'), but also produced with a different part of the tongue than /k/ (difference of 'kind'). Though areas 3 and 4 are labelled intermediately (after 1 and 2—with 2 and 3^- overlapping in 'position'; and before 5, 6), the consonants labelled by these numbers (3 and 4) are generally rather than strictly intermediate in 'position' and hence are better identified in terms of 'kind' than of 'position'. Number label 3 is used for affricates with sibilant off-glide, and for the sibilants themselves (s, š), and for spirantal y; number label 4 is used for liquids—any kind of r or l sound—and for affricates with 1 off-glide. By letter label, an affricate is S (a stop), but S at 3 is always an affricate with sibilant off-glide, and S at 4 is always an affricate with 1 off-glide (/λ/ for [dl] and /ƛ/ for voiceless [tɬ]).

Number labels and letter labels seem awkward only in first exposition; they prove convenient in making notations from phonemic systems, and in comparing parts of one system with parts of another. For example, the next six paragraphs give a survey of all the consonants found in the languages listed above (a).

We find S and N at 1 = /p m/ (Hidatsa, Shawnee, Totonac); also, F and N at 1 = /ɸ m/ (Waiwai); also S and F and N at 1 = /p ɸ m/ (Maori). When a language shows a distinction between bilabial consonants and labio-dental consonants, the latter are numbered 1^+ and said to be 'in' area 1 rather than 'at' 1; the term 'at' is restricted to two or more matched labial consonants which are not only 'in' area 1 but are produced with the same articulators in the area in question, as with both lips, /p m/. Thus, Havasupai shows three LC 'in' 1 = /p m, f/; of these, the S and N are said to be 'at' 1 = /p m/, while the F, though 'in' 1, is (more precisely) 'at' 1^+ = /f/.

We find S and N at 2 = /t n/ (Maori, Hidatsa, Totonac). The Totonac N at 2 is followed in notation by a parenthesis to show velar allophones which occur in a non-2 area; thus, 2 (∼ 5, 5^+) = /n ([n ∼ ŋ ∼ ŋ̟])/, or more briefly /n (∼ [ŋ])/. (In the general case for all languages, allophones of a phoneme remain in a given numbered area; the alternation between alveolar [n] and velar [ŋ] proves to be the only wide spread instance of allophones appearing in more than one of our numbered areas.) When a language shows a distinction between alveolar and interdental consonants, the latter are numbered 2^- and are said to be 'in' area 2 but 'at' 2^-. Thus, Shawnee shows three LC in 2 = /θ, t n/; of these the F is at 2^- = /θ/, while the S and N are at 2 = /t n/. Havasupai is in part like Shawnee, but makes an additional distinction 'in' area 2—an alveolar retroflex S = /ʈ/ which is numbered 2^+. Altogether, Havasupai 2 = /θ, t n, ʈ/; of these, the F is said to be at 2^- = /θ/, the first S and N at 2 = /t n/, the second S at 2^+ = /ʈ/. Waiwai makes another kind of distinction 'in' area 2—an alveolar palatalized N = /ñ/ which is numbered 2 prime ($2'$). Altogether, the Waiwai LC in area 2 are /t n, ñ/; of these the S and first N are at 2 = /t n/, while the second N is at $2'$ = /ñ/.

We find languages with no consonants in 3 (Maori), as well as languages with more LC in 3 than in any other area. In our restricted sample, we find S and F at 3^{-} = /c s/ (Havasupai); also at 3 = /č š/ (Shawnee; and Waiwai where, beside S and F at 3, there is another F at 3^{-} = /s/); also at both 3^{-} and 3 = /c s, č š/ (Totonac).

Number label 4, without superscript, is used for l liquids; however for r liquids, one or another superscript number (suggesting 'position') is written after 4—4^{2} for front /r/, 4^{5} for back /ɽ/— with the superscript doubled for trilled /r̃/ (4^{22}). We find L at 4^{2} = /r/ (Maori); also at 4 = /l/ (Shawnee; and Totonac, with voiceless /ɬ/ as well as voiced /l/; also at both 4^{2} and 4 = /r, l/ (Havasupai); also at both 4^{2} and 4^{22} = /r, r̃/ (Waiwai). Our sample shows one language with no consonants in 4 (Hidatsa), but since the allophones of /n/ include [r], we could draw attention to this in the notation for Hidatsa by writing N at 2 (~ L at 4^{2}).

We find S at 5 = /k/ (Shawnee and Waiwai); also S at both 5 and 5^{+} = /k, q/ (Havasupai and Totonac); also S and N at 5 = /k ŋ/ (Maori); also S and F at 5 = /k x/ (Hidatsa).

We find S and F at 6 = /ʔ h/ (Hidatsa, Havasupai); also F alone at 6 = /h/ (Maori, Waiwai, Totonac); also S alone at 6 = /ʔ/ (Shawnee, where [h] is an allophone of /ʔ/ but [ʔ] is the more conspicuous allophone).

(c) It is now possible, from information given above under (b), to compute type formulas which will show proportional matchings—i.e., the LC occurrence of stop and fricative (S-F) in two or more numbered areas; and likewise of stop and nasal (S-N); and finally, in our restricted sample, of stop and fricative and nasal (S-F-N). Each language exemplified turns out to have a different consonant type, in spite of the many partial similarities shown in the survey above. The most elaborate of these formulas is the Havasupai type: Quadruple S-F (in 1, 2, at 3^{-} 6) + Double S-N (at 1, 2) + Double S-F-N (in 1, 2) = /p f, t θ, c s, ʔ h/ + /p m, t n/ + /p f m, t θ n/. In consonant type, Shawnee and Totonac are almost identical, and indeed would appear wholly identical if the information on occurrences in numbered areas were withheld; this information is given parenthetically in the type formulas. The Shawnee type is Double S-N (at 1, 2) + Double S-F (in 2, at 3) = /p m, t n/ + /t θ, č š/. The Totonac type is Double S-N (at 1, 2) + Double S-F (at 3^{-}, 3) = /p m, t n/ + /c s, č š/.

The preceding types show compound formulas; compare now simple formulas—for the Maori type: Triple S-N (at 1, 2, 5) = /p m, t n, k ŋ/; and for the Hidatsa type: Triple S-F (in 3, at 5, 6) = /c š, k x, ʔ h/.

Waiwai does not show matching of F-N, for example, in more than one numbered area (F-N only in 1 = /ɸ m/; S-N only in 2 = /t n/; S-F only in 3 = /č š/)—this lack of double or more matchings, though single matchings occur, is termed ZERO PROPORTIONAL MATCHING.

II. Turning now from the selected fraction of our sample—languages without consonant additive components in which the sum of linear consonants (LC) is necessarily the same as the sum of all unit consonants (C)—we find that the range of diversity in consonant types for the whole sample is about the same as that given above. There are other instances of simple formula, besides those given for Maori and Hidatsa. For example, both the Winnebago and the Chatino consonant type is Double S-N (at 1, 2) = /p m, t n/.

This simple type-formula—Double S-N (at 1, 2)—is, as noted, shared by Winnebago and Chatino. Though these two languages show the same consonant type (based on LC), they differ in consonant subtype (based on AC). In Winnebago, one additive component (AC), *voicing*, combined with S, yields additional unit phonemes—/b ǯ g/; in

Chatino two AC, *voicing* and *prenasalization*, combined with S, also yield additional unit phonemes—/b d g/ and /ᵐb ⁿd ᵑg/. If (instead of relegating this information to subtypes based on AC) one attempted to incorporate the additional unit phonemes into one consonant type—if voiced stops were matched against voiceless stops—then Winnebago would not show the same consonant type as Chatino. Winnebago would show fewer matchings than Chatino, would be classified as a different consonant type. But we still need to justify our preference for classifying Winnebago and Chatino as sharing one consonant type: Double S-N (at 1, 2), with subtype classification of one AC for Winnebago (*voicing*) and two AC for Chatino (*voicing* and *prenasalization*).

By listing AC for subtypes rather than matching the additional unit phonemes against the same 'kind', we can hope to arrive at identical or similar consonant types for single languages which are phonemicized differently. Hidatsa was phonemicized as having lenis and fortis consonants by Harris and Voegelin and as having a single series (with consonant clusters instead of fortis unit phonemes) by Robinett; in both phonemicizations, the consonant type is the same, as cited; the difference in phonemicization is reflected in subtype—with an AC of *fortisness* in the earlier phonemicization; without this AC in the later consonant cluster solution of the phonemic problem. (This is not an isolated instance; not infrequently phonemic systems seem amenable to alternative phonemicizations.)

Often enough languages which show the same or similar consonant types (based on LC) will show some difference in subtype (based on AC). Thus, the compound formula for the Havasupai type is similar to that for the Choctaw type. The Havasupai type is: Quadruple S-F + Double S-N + Double S-F-N (disregarding occurrences of LC in numbered areas). The Choctaw type is: Quintuple S-F + Double S-N + Double S-F-N. However, in subtype, based on AC, the two languages differ, since Choctaw includes a consonant AC (*voicing*), while Havasupai does not.

But among languages in the same language family, it often seems as though there were a diminishing correlation between consonant type and subtype, as one goes from more closely to less closely related languages. The more distantly related language may differ in consonant type from the type shared by a closely related group of languages, but still retain the same consonant subtype as the latter—i.e., the same shared AC. Thus, in subtype, the genuinely Athapaskan Navaho and Chipewyan, as well as Tlingit—at best a distant relative in an extended Athapaskan family—all share two consonant AC, *voicing* and *glottalization*. However, for consonant type it is only Chipewyan and Navaho that are closely similar in compound type-formula: Seven-fold S-F (for Chipewyan) and Six-fold S-F (for Navaho A) + Quadruple S-NL (for both) + Double S-F-NL (for both). Tlingit, on the other hand, shows a simple type-formula: Seven-fold S-F, with no proportional matchings for S-NL, nor for S-F-NL.

Though no additive components appear in the consonant systems of some languages (those discussed under **I**, for example), all consonant systems with a large number of unit phoneme consonants (C) do include one or more AC. Our published sample ranges from 9 C to 15 C for languages without consonant AC; from 16 C to 49 C for languages with one or two AC, as *voicing* and/or *glottalization*, *implosive preglottalization*, *prenasalization*, *aspiration*, *palatalization*.

Linear consonants are all isolable—producible in isolation. Some of the AC listed above are also isolable, as *prenasalization*; others are not, as *voicing*. But all AC, whether isolable or not, are combinable with S or F or NL, and in such combination are counted as additional unit phonemes.

In analyzing the LC and AC of a given

language, the procedure is the other way around, of course. One begins with a full inventory of all unit phoneme consonants (C), and reduces these to a fewer number of linear consonants (LC) whenever it is possible to extract an additive component (AC); but this is not always possible. The AC extracted is not some predetermined phonetic feature, as *voicing* or *palatalization*, or one of the other AC listed above. In Totonac, for example, there is a voiceless /ɬ/ and voiced /l/, and yet it is not possible, in this system, to extract *voicing* as an AC because /ɬ/ and /l/ both fall in area 4, and *voicing* does not distinguish any other consonant in any other numbered area. But in other systems, as in Choctaw, *voicing* distinguishes /b/ from /p/ in area 1, and /g/ from /k/ in area 5; in the Choctaw system *voicing* may be extracted as an AC. The criterion for extracting AC is the same as the criterion for matching LC, and of course all AC could be restated as instances of proportional matching—with exactly the same 'kind' of consonant.

In an earlier attempt to focus on such same 'kind' matching (UCPL 10.27–45, 1954), I was forever noting asymmetries—holes in the pattern. Had I then treated Choctaw, I would have noted that the matching between /p t k/ and /b g/ is not symmetrical because Choctaw lacks a voiced t. But with *voicing* extracted as an AC, one merely notes that in Choctaw *voicing* combines with S at area 1 and area 5 yielding two additional stops, /b g/. The latter way of putting it is completely convertible with the earlier way. But attention is shifted from matching between same 'kind' of consonants—as between stops, with all the fuss and bother over asymmetry—to LC matching between different 'kind' of consonants, as S and F and N and L. The new result is that all symmetry is specified, but in two ways: (a) for the classification of consonant types—in terms of matching between different 'kinds' of consonants, as S and F, in the same numbered area; (b) for subtypes—in terms af AC, as 'voicing', for two or more consonants of the same 'kind', as stops.

III. The third statement is concerned with linear vowels (LV), which are involved in the construction of vowel types. All linear vowels are isolable (producible in isolation); variety among the linear vowels in any given system is produced by difference in tongue positions, at the very least; and at most, in addition, by spreading or rounding of the lips. Our published sample shows languages with 3 LV (Totonac, Amuesha), 4 LV (Shawnee, Navaho), 5 LV (Maori, Hidatsa, Winnebago, Havasupai, Chatino, Aguacatec), 6 LV (Choctaw, Tarascan, Chontal, Waiwai, Witoto, Chipewyan), 8 LV (Tlingit), and 10 LV (Yagua).

To cite the number of linear vowels in a vocalic system gives enough information to imply the vowel type (so far as our sample goes) for 3 LV, 4 LV and 5 LV. Thus, both vocalic systems cited for 3 LV make a contrast between front vowel (F) and back vowel (B) at a higher tongue height than another vowel; this other vowel is neutral or non-contrastive (N) in respect to frontness and backness. In formula, (FB) = /i u/ (also transcribed as /i o/), and N = /a/; altogether the type formula is (FB) + N = /i u/a/. So also, both systems cited for 4 LV make one contrast between (FB) = /i o/ at high tongue height, and another contrast between (FB) = /e a/ at a lower tongue height; hence the whole type formula is 2 (FB) = /i o/e a/. So also, the half dozen vocalic systems cited for 5 LV make a contrast between (FB) = /i u/ at high tongue height, between (FB) = /e o/ at mid tongue height; but no front-back contrast appears at low tongue height—N = /a/; this is expressed in the type formula 2 (FB) + N = /i u/e o/a/.

However, the sum total of LV does not imply a single vowel type when the total exceeds 5 LV. Thus, languages cited for 6 LV show two fundamentally different

vowel types. In the Chipewyan type, three front-back contrasts are made at high, mid, and low tongue heights; hence, 3 (FB) = /i u/e o/ɛ a/. In the other languages cited for 6 LV, the type is less symmetrical. A trinary contrast appears either at high tongue height, (FCB) = /i i u/ (in Chontal, Waiwai and Witoto), or at mid tongue height, (FCB) = /e ə o/ (in Choctaw and Tarascan), with a binary front-back contrast (mid (FB) = /e o/, when the trinary contrast is high; but high (FB) = /i u/, when the trinary contrast is mid), and with a non-contrastive vowel, N = /a/, at low tongue height. These two systems are represented in type formula as (FCB) + (FB) + N = /i i u/e o/ a/; and as (FB) + (FCB) + N = /i u/e ə o/a/.

In all of the vowel systems cited, rounding is a redundant correlate of back vowels, and front vowels are uniformly unrounded. In Yagua, however, one high front vowel is unrounded, another rounded (FF° = /i ü/); and the same distinctive rounding distinguishes two other front vowels at lower mid tongue position (FF° = /ɛ ö/). Each pair of front vowels contrasts with a back vowel. This half of the Yagua vowel system is represented in formula as 2(FF° B) = /iü u/ɛö o/. The remaining half of the same system shows a binary front-back contrast at mid tongue height (FB = /e o/) and at low tongue height (FB = /ə a/); hence, + 2(FB) = /e o/ə a/.

IV. Two languages may share the same vowel type, based on linear vowels (LV), but differ in subtype. Vowel subtype is determined by additive components (AC) which are combinable with vowels but are not isolable (not producible in isolation)—*nasalization*, *glottalization*, *length*, *stress*, *syllabic tones*. Except for those languages in the sample which include syllabic tones in their phonemic inventories, one or two AC—not more than two—are combinable with some or all LV. The single AC involved is *length* (Maori, Shawnee, Aguacatec), or *stress* (Chontal, Tarascan, Waiwai, Witoto); double AC involved are *length* + *stress* (Hidatsa, Havasupai), *length* + *glottalization* (Totonac, Amuesha), *length* + *nasalization* (Choctaw).

Perhaps the most striking instance of subtype differences between any two languages is that of Shawnee and Navaho which, as noted under **III,** share the same vowel type—2(FB) = /i o/e a/. In Shawnee, a single AC, *length*, is combinable with all LV in the system. For typological purposes, it is immaterial whether Shawnee length is phonemicized as LV + /·/ = /i· o· e· a·/, or phonemicized as LV + identical LV = /ii oo ee aa/ (cluster solution). In either case, the number of syllabics is doubled, from four to eight—from four syllabics, counting LV alone, to eight syllabics, when the AC *length* is combined with LV. In Navaho six AC (*nasalization*, *length*, four syllabic tones—*high*, *low*, *low-high*, *high-low*) are combinable not only with LV in the system, but some of the AC—some of the syllabic tones—are combinable with a non-syllabic /n/ which is then transformed to a syllabic, yielding a total of 50 (or more) syllabics (depending on how the Navaho vowel syllable system is phonemicized).

Most information on semivowels is tentatively included in statement **IV,** as relevant to the construction of a second vowel subtype, having to do with diphthongs, or more generally, with clustering of vocalic or open contact phonemes, including semivowels (but not /h/ which is typologized as a consonant). For most languages, both /w/ and /y/ are added to the sum of syllabics; or /w/ alone (Maori, Winnebago); or neither. Spirantal /y/ is counted as a linear consonant, F in area 3 (Navaho, Chipewyan, Tlingit); so also /w/, F in area 1 (Chontal).

V. It is traditional in linguistics to speak of a three vowel system, of a four vowel system, or of that most common system of all, a five vowel system—referring, of course,

to the isolable vowels irrespective of different vocalic components, as length or nasalization, which may or may not be involved in two different vowel systems being discussed. This traditional way of talking is merely formalized in vowel types and such summary statements as 3 LV, 4 LV, and 5 LV. For consonant systems, however, there is no traditional analogue for three, or four, or five vowel systems. There is accordingly no ready parallel or available summary reference to the general nature of a consonant system. The formal analogy to 3 LV or 5 LV is, of course 9 LC or 20 LC.

And the general relationship between the linear consonants in a given language and the linear vowels in the same language can be expressed as a ratio. The range of such ratios appears to be fairly narrow, possibly because linear consonants always outnumber linear vowels, at least 2:1 (Maori, Hidatsa, Winnebago, Choctaw, Waiwai, Chatino, Witoto, Tlingit, with a little less for Yagua and a little more for Tarascan). LC also outnumber LV in the ratio 3:1 (Shawnee, Havasupai, Chontal, Aguacatec, Chipewyan), and 4:1 (Navaho), and 5:1 (Amuesha).

VI. It is not unusual to speak of a language as having two series of consonants, as voiced and voiceless, or to say that another language has only one series, or that some Indian languages on the West Coast have four series of stops. Nor is it unusual to speak of a vowel system as having nasalized vowels. This common usage is again merely formalized in consonant subtypes and in vowel subtypes, and by such summary statements as zero AC, one AC, or two AC; or for that matter, for vowels, six AC.

A general relationship between consonant subtype and vowel subtype is expressible as a ratio of consonant AC to vowel AC. Our sample is conspicuous for its large number of languages lacking consonant AC; it includes no instance of zero vowel AC. (But languages lacking vowel additive components are known outside the sample.) Though consonant AC may outnumber vowel AC, they do so at most 2:1 (Tlingit, Tarascan, Chontal, Witoto). Consonant and vowel AC may be evenly matched, 1:1 (Winnebago, Aguacatec, Amuesha), but mostly the consonant additive components in a given language are outnumbered by the vowel additive components in the ratio of 1:2 (Choctaw), 1:3 (Navaho, Chipewyan, Chatino), 1:4 (Yagua), 0:1 (Maori, Waiwai), and 0:2 (Hidatsa, Havasupai, Totonac).

FREQUENCIES AND INVENTORIES OF PHONEMES FROM NINE LANGUAGES

JOHN YEGERLEHNER
UNIVERSITY OF ARIZONA

FLORENCE M. VOEGELIN
INDIANA UNIVERSITY

AND OTHERS

(Content **1**, below)

0. Introduction
1. Phonemic inventories by:
Bruce Biggs (Maori)
Florence M. Voegelin (Hidatsa)
G. Hubert Matthews (Winnebago)
John Yegerlehner (Shawnee)
Gordon Hirschy (Choctaw)
Jerry Hopkins (Havasupai)
Kenneth Hale (Navaho)
Kathryn C. Keller (Chontal)
Alan Wares (Tarascan)
2. Typological and frequential alignment
3. Chart of rank orders

0. A number of choices are possible in selecting an arrangement of phonemes for cross-language comparisons. Here, one possible arrangement—that presented in the paper immediately preceding this one[1]—is selected, and an attempt is made to correlate relative frequency statements with typological statements by comparing the relative frequencies both of phonemes grouped together in the typology and of groups of phonemes opposed in the typology—linear phonemes *vs.* additive component phonemes, matched consonants *vs.* unmatched ones, front vowels *vs.* back vowels, and so on.

This paper is intended to be no more than an exploration of some of the possibilities of comparisons between arrangements of phonemic inventories and the relative frequencies of the phonemes, and statistical validity is not assumed for the results obtained from comparisons between so few languages on the basis of so small a textual sample of each. The size of the sample in terms of the number of languages and the number of phonemes counted for each was determined by the languages controlled by the members of a seminar at Indiana University and the length of texts available in those languages. The sample unfortunately does not present either the desired possibility of languages for which the typological statements are identical or languages with a wide range of diversity.

The phonemic inventories of the nine languages in the sample are presented below (**1.**) in terms of C. F. Voegelin's Six Statements for a Phonemic Inventory. The relative frequencies of the phonemes in each language were established on the basis of their occurrence in a sample text of 2000 phonemes, and the frequency rank orders are presented in **3.**, below.

1. The phonemic inventories for the nine languages in the sample are given below.[2]

MAORI (NEW ZEALAND, MALAYO-POLYNESIAN)

I. 9 LC at 1; 2; 4; 5; 6 = /p ϕ m/ t-n/r/k ŋ/h/ = consonant type Triple S-N (at 1, 2, 5); with single matching of S-F-N (at 1), with unmatched F (at 6) and L (at 4[2]).

[1] C. F. Voegelin, Six Statements for a Phonemic Inventory, IJAL 23.78-84 (1957), *q.v.* for definition of terms 'linear phoneme,' 'additive component,' 'matching.'

[2] Abbreviations used in this section: B = Back, C = Central, F (as in 'FCB') = Front, N (as in 'FCB + N') = Neutral (with respect to the opposition Front *vs.* Back), LC = Linear Consonant, LV = Linear Vowel, S = Stop, F (as in 'Double S-F' or 'unmatched F') = Fricative, N (as in 'Double S-N') = Nasal, NL = Nasal-liquid, AC = Additive Component.

II. No consonant AC; hence 9 LC = all C in inventory.

III. 5 LV = /i u/e o/a/ = vowel type 2(FB) + N.

IV. One vowel AC: *length* (by cluster solution) combinable with 5 LV (5 + 5 = 10 syllabics), with /w/ = 11 syllabics.

V. Ratio of LC to LV = <2:1 for 14 linear phonemes.

VI. Ratio of consonant AC to vowel AC = 0:1.

HIDATSA (NORTHERN PLAINS, SIOUAN)

I. 10 LC at 1; 2; 3^{-}, 3; 5; 6 = /p m/t n/c, š/k x/ʔ h/ = consonant type Triple S-F (in 3, at 5, 6), Double S-NL (at 1, 2); with no single matching, with no unmatched LC.

II. No consonant AC; hence 10 LC = all C in inventory.

III. 5 LV = /i u/e o/a/ = vowel type 2(FB) + N.

IV. Two vowel AC: *length* combinable with 5 LV (5 + 5 = 10 syllabics); *stress* combinable with short and long V (10 + 10 = 20 syllabics).

V. Ratio of LC to LV = 2:1 for 15 linear phonemes.

VI. Ratio of consonant AC to vowel AC = 0:2.

WINNEBAGO (EASTERN WOODLANDS, SIOUAN)

I. 12 LC at 1; 2; 3^{-}, 3; 4^{2}; 5; 6 = /p m/t n/s, č š/r/k x/ʔ h/ = consonant type Triple S-F (at 3, 5, 6), Double S-N (at 1, 2); with no single matching, with unmatched F (at 3^{-}), L (at 4^{2}).

II. One constant AC: *voicing* at 1, 3, 5 combined with S = /b ǯ g/; at 3^{-}, 3, 5 with F = /z ž γ/; hence these 6 C + 12 LC = 18 for all C.

III. 5 LV = /i u/e o/a/ = vowel type 2(FB) + N.

IV. One vowel AC: *nasalization* combinable with /i u/ and /a/ (not with /e o/); hence 3 + 5 LV = 8 syllabics, with /w/ = 9 syllabics.

V. Ratio of LC to LV = >2:1 for 17 linear phonemes.

VI. Ratio of consonant AC to vowel AC = 1:1.

SHAWNEE (EASTERN WOODLANDS, ALGONQUIAN)

I. 11 LC at 1; 2^{-}, 2; 3; 4; 5; 6 = /p m/θ, t n/č š/l/k/ʔ (~[h])/ = consonant type Double S-F (in 2, at 3), Double S-N (at 1, 2); with single matching of S-F-N (in 2)—not to mention and not to count allophonic matching of S ~ F at 6; with unmatched S (at 5) and L (at 4).

II. No consonant AC; hence 11 LC = all C in inventory.

III. 4 LV = /i o/e a/ = vowel type 2(FB).

IV. One vowel AC: *length* (by cluster solution) combinable with 4 LV (4 + 4 = 8 syllabics), with /w y/ = 10 syllabics.

V. Ratio of LC to LV = <3:1 for 15 linear phonemes.

VI. Ratio of consonant AC to vowel AC = 0:1.

CHOCTAW (SOUTHEAST, MUSKOGEAN)

I. 14 LC at 1, 1^{+}; 2^{-}, 2; 3^{-}, 3; 4; 5; 6 = /p m, f/θ, t n/s, č š/l/k x/ʔ h/ = consonant type Quintuple S-F (in 1, 2, at 3, 5, 6), Double S-N (at 1, 2), Double S-F-N (in 1, 2); with unmatched F (at 3^{-}) and L (at 4).

II. One consonant AC: *voicing* at 1, 5 combined with S = /b g/; hence these 2 C+14 LC = 16 for all C.

III. 6 LV = /i u/e ə o/a/ = vowel type (FB) + (FCB) + N.

IV. Two vowel AC: *nasalization* and *length* combinable with 6 LV (6 + 6 + 6 + 6 = 24 syllabics), with /w y/ = 26 syllabics.

V. Ratio of LC to LV = >2:1 for 20 linear phonemes.

VI. Ratio of consonant AC to vowel AC = 1:2.

HAVASUPAI (SOUTHWEST, YUMAN)

I. 15 LC at 1, 1^{+}; 2^{-}, 2, 2^{+}; 3^{-}; 4^{2}, 4; 5, 5^{+}; 6 = /p m, f/θ, t n,ṭ/c s/r, l/k, q/ʔ h/ =

consonant type Quadruple S-F (in 1, 2, at 3^{-}, 6), Double S-N (at 1, 2), Double S-F-N (in 1, 2); with unmatched S (at 2^{+}, 5, 5^{+}) and L (at 4^{2}, 4).

II. No consonant AC; hence 15 LC = all C in inventory.

III. 5 LV = /i u/e o/a/ = vowel type 2(FB) + N.

IV. Two vowel AC: *stress* and *length* which combine with 5 LV and each other (5 + 5 + 5 + 5 = 20 syllabics), with /w y/ = 22.

V. Ratio of LC to LV = 3:1 for 20 linear phonemes.

VI. Ratio of consonant AC to vowel AC = 0:2.

NAVAHO (SOUTHWEST, ATHAPASKAN)

I. 16 LC at 1; 2; 3^{-}, 3; 4; 5^{0}, 5; 6 = /(p*) m/t n/ c s, č š y/ x ɬ l/k^{w} x^{w}, k x/ʔ/ = consonant type Quintuple S-F (at 3^{-}, 3, 4, 5^{0}, 5), Quadruple S-NL (1, 2, 3, 4), Double S-F-NL (at 3, 4); with no single matching, with unmatched S (at 6).

II. Two consonant AC: *voicing* at 1, 2, 3^{-}, 3, 4, 5 combined with S = /b d ʒ ǯ λ g/, and at 3^{-}, 3, 5^{0}, 5 with F = /z ž γ^{w}ɣ/; *glottalization* at 2, 3^{-}, 3, 4, 5 with S = /t' c' č' x' k'/ and at 1, 2, 3 with NL = /m' n' y'/; hence these 18 C + 16 LC = 34 for all C.

III. 4 LV = /i o/e a/ = vowel type 2(FB).

IV. Six vowel AC: *nasalization* combined with 4 LV (4 + 4 = 8 syllabics); *length* by cluster solution (8 + 8 = 16 syllabics); plus either *low* or *high tone* for each (16 + 16 = 32 syllabics), plus *high-low* with any long syllabic (8 + 32 = 40), plus *low-high* with any long syllabic (8 + 40 = 48), plus *high tone*, *high-low* or *low-high* with /n/ for syllabic /ń ńn̂ ǹń/ (48 + 3 = 52 syllabics).

V. Ratio of LC to LV = 4:1 for 20 linear phonemes.

VI. Ratio of consonant AC to vowel AC = 1:3.

CHONTAL (MEXICO, MAYAN)

I. 16 LC at 1, 1^{+}; 2; 3^{-}, 3; 4^{2}, 4^{22}, 4; 5; 6 = /p m, w/t n/c s, č š, y/r, r̃, l/k/ʔ h/ = consonant type Triple S-F (at 3^{-}, 3, 6), Triple S-NL (at 1, 2, 3); with single matching of S-F-L (at 3), with unmatched S (at 5) and L (at 1^{+}, 4^{2}, 4^{22}, 4).

II. Two consonant AC: *glottalization* at 1, 2, 3^{-}, 3, 5 combined with S = /p' t' c' č' k'/; *voicing* at 1, 2, 5 with S = /b d g/; hence these 8 C + 16 LC = 24 for all C.

III. 6 LV = /i i u/e o/a/ = vowel type (FCB) + (FB) + N.

IV. One vowel AC: *stress* combinable with 6 LV (6 + 6 = 12 syllabics).

V. Ratio of LC to LV = <3:1 for 22 linear phonemes.

VI. Ratio of consonant AC to vowel AC = 2:1.

TARASCAN (MEXICO, UNCLASSIFIED)

I. 15 LC at 1, 1^{+}; 2; 3^{-}, 3; 4^{2}, 4^{2+}, 4; 5; 6 = /p m, f/t n/c s, č š/r, ɾ, l/k ŋ /h/ = consonant type Triple S-F (in 1, at 3^{-}, 3), Triple S-N (at 1, 2, 5); with single matching of S-F-N (in 1), with unmatched L (at 4^{2}, 4^{2+}, 4) and F (at 6).

II. Two consonant AC: *aspiration* at 1, 2, 3^{-}, 3, 5 combined with S = /p^{h}, t^{h}, c^{h} $č^{h}$ k^{h}/; *voicing* at 1, 2, 5 with S = /b d g/; hence these 8 C + 15 LC = 23 for all C.

III. 6 LV = /i u/e ə o/a/ = vowel type (FB) + (FCB) + N.

IV. One vowel AC: *stress* combined with 6 LV (6 + 6 = 12 syllabics).

V. Ratio of LC to LV = $2\frac{1}{2}$:1 for 21 linear phonemes.

VI. Ratio of consonant AC to vowel AC = 2:1.

2.0. The arrangement of phonemes into series (lines) and areas (columns) and the weighting in the typological statements of linear (as opposed to additive component) phonemes and of matched linear phonemes (as opposed to unmatched linear phonemes) were compared with the relative frequency arrangements—rank orders—in several

ways. Comparisons of frequencies were made within each language and across languages between linear consonants and their matched linear consonants in different series, between matched linear consonants and unmatched ones, between areas in each series, between areas across series, between linear consonants and vowels and their additive component counterparts, and between vowels in distinctive positions. The results of these comparisons are given below in the order of the six statements to which they apply.

2.1.1. The frequencies of pairs of matched linear phonemes within a given area were compared. Thus the relative frequencies of the linear stops and their corresponding matched fricatives within an area were compared, as were the relative frequencies of matched stops and nasal-liquids and of matched nasal-liquids and fricatives.

In Chontal, Hidatsa, and Navaho, linear fricatives were more frequent than their corresponding matched stops, except that in Chontal /č/ was more frequent than /š/ (as it was in four of the six languages in which both occurred), that in Hidatsa /k/ was more frequent than /x/ (as it was in three of the four languages in which both occurred), and that in Navaho /x^{w}/ and /k^{w}/ both occurred only once in the sample text. In Shawnee, Choctaw, Winnebago, Tarascan, and Havasupai, linear stops were more frequent than their corresponding matched fricatives, except that in Choctaw and Winnebago /h/ was more frequent than /ʔ/ (as it was in all of the five languages in which both occurred), that in Tarascan /s/ was more frequent than /c/ (as it was in four of the five languages in which both occurred), and that in Havasupai /f/ was more frequent than /p/.

Of the four languages in our sample having both a stop (/p/) and a fricative (/f/ or /ɸ/) in area 1, stop was more frequent than fricative in three. In area 2, three languages out of three had stop (/t/) more frequent than fricative (/θ/). At 3, fricative (/s/) was more frequent than stop (/c/) in three out of four cases; also, in one case of skew matching in area 3 between /š/ and /c/, the former was more frequent than the latter. At 3, stop (/č/) was more frequent than fricative (/š/) in four out of six cases. In area 5, stop (/k/) was more frequent than fricative (/x/) in three out of four cases. In area 6, fricative (/h/) was more frequent than stop (/ʔ/) in five out of five cases.

In Chontal, Tarascan, and Maori, linear stops were more frequent than their matched nasal-liquids, except that in the former two, /n/ was more frequent than /t/ (as it was in five of the nine languages in which both occurred) and that in Maori, /m/ was more frequent than /p/ (as it was in six of the eight languages in which both occurred). In Shawnee, Choctaw, and Havasupai, /m/ was more frequent than /p/ while /t/ was more frequent than /n/. In Navaho, Hidatsa, and Winnebago, all linear nasal-liquids were more frequent than matched stops.

Comparison of the relative frequencies of matched linear stops and nasal-liquids showed that in area 1, /m/ was more frequent than /p/ in six out of eight cases; in area 2, /n/ was more frequent than /t/ in five out of nine cases; at 3, /č/ was more frequent than /y/ in one language, but /y/ was more frequent than /č/ in a second; in area 5, /k/ was more frequent than /ŋ/ in the two cases included in our sample.

Fricative-Nasal Liquid matching occurred in the sample only as part of Stop-Fricative-Nasal Liquid matching. In nine of the ten cases of such matching, the nasal liquid was more frequent than the fricative, i.e., /m/ was more frequent than /f/ or /ɸ/; /n/ was more frequent than /θ/; /l/ was more frequent than /ɬ/; and in one of the two cases of matching of /y/ and /š/, /y/ was more frequent than /š/. The frequency relationships of the stops and fricatives in these matchings are included in

the discussion of Stop-Fricative matchings above.

2.1.2. For each language, the frequencies of unmatched linear consonants were compared with those of matched linear consonants in the same series. Fourteen of the twenty-one unmatched linear consonants in the whole sample were either more frequent than any of the matched members of the same series in the language in which they occurred, or less frequent than any of the matched members of the same series; the remaining seven unmatched consonants were more frequent than some matched consonants in the same series, but less frequent than others. In Havasupai, unmatched /k/ was the most frequent stop; unmatched /q/ and /t̩/ were the least frequent stops; and unmatched /l/ and /r/ were the least frequent nasal-liquids. In Shawnee, unmatched /k/ was the most frequent stop, and unmatched /l/ the most frequent nasal-liquid. In Maori, unmatched /h/ was more frequent than the one matched fricative, and unmatched /r/ was more frequent than the matched nasal-liquids. In Navaho, unmatched /ʔ/ was more frequent than any matched stop. In Choctaw, unmatched /l/ occurred the same number of times as matched /n/, which was more frequent than the other matched nasal-liquid. In Chontal, unmatched /w/, /r/ and /r̆/ were less frequent than any of the matched nasal-liquids, but unmatched /l/ occurred more frequently than two of the matched nasal-liquids, but less frequently than one; unmatched /k/ occurred more frequently than three of the matched stops, but less frequently than two. In Tarascan, unmatched /h/ occurred more frequently than two of the matched fricatives, but less frequently than one; unmatched /l/ occurred less frequently than any of the matched nasal-liquids, but unmatched /r/ occurred more frequently than two of the matched nasal-liquids, but less frequently than one, and unmatched /ɾ/ occurred more frequently than one, but less frequently than two. In Winnebago, unmatched /s/ was less frequent than two of the matched fricatives, but more frequent than one, and unmatched /r/ was more frequent than one of the matched nasal-liquids, but less frequent than the other.

2.1.3. The frequency ranks for linear consonants within each series for each language (as, Maori: stops: /t/ is more frequent than /k/, which is more frequent than /p/) are not presented here since they may be readily seen from the chart (**3.**, below), but the following generalizations can be made when the rank orders of consonants within each series are compared across languages. Eight of the nine languages have the stops /p/, /t/, and /k/; in all of these languages /k/ and /t/ (areas 5 and 2) were more frequent than /p/ (area 1), or, in one (Winnebago) /t/ occurred the same number of times as /p/. The phoneme /k/ was more frequent than /t/ in six languages, and /t/ was more frequent than /k/ in the other three. For the sample as a whole, /k/ was the most frequent stop in five of the languages, the second most frequent in three (either /t/ or /ʔ/ was more frequent), and third in one (in which both /t/ and /ʔ/ were more frequent). In the eight languages which had five or more stops, including one or more affricate, no affricate was higher in rank among stops than third, i.e., all affricates were less frequent than at least two other stops. In the fricative series, /h/ was most frequent in six of the languages, and possibly in a seventh, Navaho, in which /h/ and /x/ were counted together as /x/, which ranked as the most frequent fricative. In Tarascan, /h/ was second in frequency among fricatives (/s/ being first). In the three languages (other than Navaho, discussed above) which have /x/ and the three languages which have /θ/, /x/ and /θ/ were less frequent than the fricative in area 3 (/s/ or /š/). In the nasal-liquid series, /n/ was more frequent than /m/ in eight of the nine languages, and /m/ was more frequent than, or as

frequent as, /ŋ/ in the two languages which had /ŋ/. Among the relative ranks of liquids in area 4 no consistencies between languages were found, except that in the two languages which had liquids both at 4^2 and at 4^{2+} or 4^{22}, the liquid at 4^2 (/r/) was more frequent than that at 4^{2+} (/ɽ/) or at 4^{22} (/r̃/).

2.1.4. When, for each language the occurrences of all the linear consonants were added together and the total frequencies of the areas were ranked, no two languages were found to have all areas in the same rank order. The relative ranks of the six areas are summarized below for all nine languages. (There were no area 3 consonants in one language, Maori, and no area 4 consonants in another, Hidatsa.)

In five languages the total number of occurrences of linear consonants in area 2 was greater than the total number of occurrences of linear consonants in any other area; area 2 was second in rank in two languages, and third in the remaining two. Area 6 was first in rank in three languages, second in one, third in two, but fourth, fifth and sixth, respectively, in the other three languages. Area 1 was first in rank in one language (Havasupai) and third in one language, but fourth in four and last (= fifth or sixth) in the other three. Area 3 was second in rank in three languages, third in one, fourth in one, fifth in two, and sixth in one. Area 5 was second in rank in two languages, third in three, fourth in three, and fifth in one. Area 4 was second in rank in one language, fourth in one, fifth in two, and last in four.

2.2. The frequencies of linear consonants were next compared with those of their corresponding additive component consonants, and frequencies between sets of additive component consonants in the same language were also compared. In Choctaw and Tarascan all linear consonants were more frequent than their corresponding AC consonants, i.e., in Choctaw linear stops were more frequent than corresponding voiced stops, and in Tarascan all linear stops were more frequent than corresponding voiced stops, which in turn were more frequent than corresponding aspirated stops. Also, in Chontal all linear stops were more frequent than corresponding AC stops (i.e., linear were more frequent than glottalized, which were more frequent than voiced, except in area 1, where voiced, but not glottalized, was more frequent than linear (/b/ more frequent than /p/, but /p/ more frequent than /p'/). In Navaho, linear fricatives and nasal-liquids were more frequent than their corresponding AC fricatives and nasal-liquids (voiced and glottalized, respectively), but linear stops were more frequent than AC stops only in area 3⁻, where linear was more frequent than either voiced or glottalized, and in area 5 where linear was more frequent than glottalized but not more frequent than voiced; elsewhere, linear stops were less frequent than either of their corresponding AC stops. All Navaho voiced stops were more frequent than corresponding glottalized stops. In Winnebago, linear consonants were equally divided between being more frequent than and less frequent than their corresponding AC consonants; one linear stop was more frequent than its voiced correspondent but two were less frequent; two linear fricatives were more frequent than their AC counterparts, but one was less frequent.

2.3. This section and the following one are devoted to relative frequencies of vowel phonemes. In this section relative frequencies are compared for linear vowels. The typological formulae for the linear vowels of all of the languages of our sample include reference to two degrees of contrastive tongue height in which tongue position at a given height (front *vs.* central *vs.* back) is contrastive. In some of the languages of the sample the phoneme /a/ is neutral (N) with respect to the opposition front *vs.* back. Frequency comparisons were made in terms of the various intersections of these dimensions in the lan-

guages of our sample. Thus the frequencies of higher front vowel (/i/) and lower front vowel (/e/) were compared against the frequencies, respectively, of higher back vowel (/u/) and lower back vowel (which in seven of the nine languages of our sample—these seven being 5- and 6-vowel systems formulatable as 2(FB) + N, (FB) + (FCB) + N, or (FCB) + (FB) + N—is the phoneme /o/, and which in the two 4-vowel systems of our sample—both of which are formulatable as 2(FB)—is /a/); also, the frequencies of higher front vowel and higher back vowel were compared against those of the corresponding vowels at the lower tongue height.

In five of the six languages of the sample for which the data were available,[3] comparisons among linear vowels showed that the higher front vowel was more frequent than the higher back vowel. In five of the six languages, the higher front vowel was more frequent than the lower front vowel. In five of the six, the lower back vowel was more frequent than the lower front vowel. In three of the six, the lower back vowel was more frequent than the higher back vowel.

In Havasupai and in Hidatsa, the higher front vowel and the higher back vowel were more frequent than the lower front vowel and the lower back vowel, respectively. In addition, in Hidatsa front vowels were more frequent than corresponding back vowels. In Chontal back vowels mere more frequent than corresponding front vowels.

The vowel /i/ ranked first in occurrence among linear vowels in Shawnee and ranked second in the other five languages under consideration. The vowel /e/ ranked last or next to last in five of the languages, but was most frequent among linear vowels in Winnebago. In Chontal the higher back vowel (/u/) was most frequent among linear vowels. In the other five languages the higher back vowel was either third or fourth in rank. In the 4-vowel system (Shawnee) the lower back vowel (/a/) ranked second among linear vowels. In the 5- and 6-vowel systems the lower back vowel (/o/) ranked third or below. In the five languages in which /a/ was neutral with respect to the front-back contrast, at least two other vowels were of less frequent occurrence than /a/. In three of these five languages /a/ was more frequent than any higher linear vowel.

2.4. This section compares the relative frequencies of linear vowels and their additive component counterparts (those vowel phonemes which include such components as stress, length, nasalization, or combinations of these). For the purposes of the typology, analysis of vowel length in terms of long vowels and in terms of geminate clusters are considered equivalent procedures, long vowels and geminate clusters both being treated typologically as vowels with the additive component of length. Our counts were made accordingly. Thus, an occurrence of the Shawnee sequence /aa/ was counted as one occurrence of /a·/.

Relative frequencies of linear vowels and corresponding AC vowels were compared in six languages. In Shawnee each linear vowel was more frequent than its corresponding long vowel. In Hidatsa each linear vowel was more frequent than each of its additive component counterparts (long, stressed, and long stressed). The back vowels (linear /u/ and /o/ and their corresponding AC vowels) were ranked as follows: linear was more frequent than stressed, which was more frequent than long stressed, which was more frequent than long. In Choctaw each of the linear vowels was more frequent than its AC counterparts. In Winnebago the high linear vowels (/i/ and /u/) were each more frequent than their nasalized counterparts, but Neutral /a/ was less

[3] The data for Maori and for Tarascan presented the combined frequencies of the various linear vowels and their additive component counterparts. The third language not considered in this section is Navaho, for reasons given below (2.4).

frequent than /ą/. In Chontal four of the six stressed vowels were more frequent than their linear counterparts (although /ó/ surpassed /o/ and /í/ surpassed /i/ only by two and by eight occurrences respectively). In Havasupai, four of the five linear vowels were more frequent than corresponding stressed vowels (/o/ greater than /ó/ by only seven occurrences). Stressed /é/ was more frequent than /e/ by one occurrence.

In Navaho, where the linear vowels do not occur in isolation, comparisons were made among four sets of AC vowels. Each of the members of a given set are marked in the same way with respect to the two additive components of length and nasalization. In counting the occurrences of the vowel phonemes, all vowels marked in the same manner with respect to these two additive components were counted together. Vowels were not further subdivided according to their tonal components. In Navaho, /e·/ was more frequent than /e/. Otherwise short non-nasalized vowels were more frequent than the corresponding long non-nasalized vowels; each of the long non-nasalized vowels was more frequent than the corresponding long nasalized vowel; each of these latter was more frequent than the corresponding short nasalized vowel. Each of the four sets of vowels showed the same internal frequency distribution; that is, higher front was more frequent than lower front, lower front was less frequent than lower back, lower back was more frequent than higher back, and higher back was less frequent than higher front.

2.5. For each language, the total number of occurrences of all consonants (linear and additive component) was compared with the total number of occurrences of all vowels; these comparisons are given below in terms of percentages of the total number of phonemes in each sample. In one language, Tarascan, the number of occurrences of vowels exactly equalled the number of occurrences of consonants. In two languages, Maori and Winnebago, vowels were more frequent than consonants: in Maori, 58.7% vowels and 41.3% consonants; in Winnebago, 51% V and 49% C. In the other six languages, consonants were more frequent than vowels: in Navaho, 56% C and 44% V; in Chontal, 55.8% C and 44.2% V; in Hidatsa, 52.6% C and 47.4% V; in Havasupai, 52.5% C and 47.5% V; in Shawnee, 53.5% C and 46.5% V; and in Choctaw, 50.8% C and 49.2% V.

3. Chart of rank orders.[4]

	M	Hi	W	S	Cc	Ha	N	Cn	T
i	2	3	3	1	2	3	2	9	2
e	3	16	1	3	13	21	16	29	7
a	1	1	4	2	1	2	1	11	1
o	6	18	15	11	4	18	7	19	11
u	9	7	10		8	10		1	6
ə					14				15
w	13		13	4	17	9		25	
y				14	11	6	17	22	
p	12	15	21	15	10	16		18	10
t	4	13	21	7	3	11	8	6	9
c		11				12	26	27	19
č			19	12	8		31	14	16
k	5	5	9	5	6	1	21	7	4
ʔ		6	14	10	23	19	3	5	
f					16	8			22
θ				18	20	21			
s			18		12	14	18	26	5
š		8	16	17	14		25	16	20
x		17	20		19		6		
m	11	9	17	13	9	4	30	20	11
n	10	2	6	9	7	13	5	2	3
l				6	7	17	11	13	19
r	8		8					28	8
ŋ	11								18
ɨ		10		16	22		8		

	M	Hi	W	S	Cc	Ha	N	Cn	T
e·		24		11	21		13		
a·		14		8	22		12		
o·		25		19	23		15		
ɨ		17				15		8	
é		24				20		12	
á		19				5		3	
ó		22				22		18	
ú		21				17		21	
ɨ·		12				24			
ú·		23				25			
į			11		21		23		
ą			2		17		27		
ǫ					18		30		
b			22		15		9	15	14
d							4	31	13
ǯ			13				22		
g			5		12		11		19
z			21				19		
ž			8				29		
γ			22				24		
t'							14	24	
c'							32	28	
č'							29	30	
k'							25	17	
h	7	4	7		5	7		4	12

[4] M = Maori; Hi = Hidatsa; W = Winnebago; S = Shawnee; Cc = Choctaw; Ha = Havasupai; N = Navaho; Cn = Chontal; T = Tarascan.

Ranks of phonemes which occured in one language only do not appear in the chart but were as follows: Maori /φ/ 14; Hidatsa /u·/ 26, /é·/ 24, /á·/ 20, /ó·/ 23; Winnebago /ɥ/ 12; Havasupai /ṭ/ 25, /q/ 23; Navaho /kʷ/ 33, /ɬ/ 10, /xʷ/ 33, /ę/ 26, /i·/ 20, /ę·/ 28, /ą·/ 25, /ǫ·/ 33, /ǯ/ 31, /ʎ/ 30; Chontal /ɨ/ 10, /ɬ/ 23, /p'/ 29; Tarascan /ɽ/ 17, /pʰ/ 23, /tʰ/ 21, /čʰ/ 23, /kʰ/ 21.

For Navaho, a given vowel symbol stands for a set of vowels marked alike with respect to the additive components of length and nasalization, regardless of which tone component is present.

A STATISTICAL STUDY OF CONSONANTS IN NEW WORLD LANGUAGES II: DATA

JOE E. PIERCE
ANKARA, TURKEY

2.3. Types
2.4. Frequencies

2.3. The purpose of this study was to investigate types, but the procedure in the preceding section gave a model that was typical for all the languages in the sample and therefore couldn't possibly be typical of any one type as opposed to another. However, the reason for doing this will become clear as the discussion proceeds.

As stated earlier, C. F. Voegelin categorized phonemic systems into languages which have one, two, three, or four stop-series,[9] and this work is a statistical test of the validity of such a grouping. The part of the investigation described in this section of the paper is an examination of the sample for things that would correlate with one type but not with another. This would show which things are characteristics of each type.

The first step in this correlation was to check and see if there was any difference in gross number of consonants per language in the four types.

Among the 176 languages in the sample, 66 had a single series of stops, i.e. these languages are then classified as type I. The maximum number of consonant phonemes for this type was 28 (Yuma),[10] a minimum of nine (Wichita),[11] and an average of 14.5 consonants per language.

67 languages in the sample had two series of stops, i.e. they were classified as type II languages. The maximum number of consonant phonemes for this type was 33 (Bella Coola)[12] a minimum of thirteen (San Francisco Solano)[13] and an average of 20.7 consonants per language.

38 languages in the sample had three series of stops, i.e. they were classified as type III. The maximum number of consonant phonemes for this type was 46 (Haida)[14] a minimum of eighteen (Maidu)[15] and an average of 29.3 consonants per language.

Five of the languages in the sample had four series of stops, i.e. they were classified as type IV. The maximum number of consonant phonemes for this type was 36 (Salinan)[16] a minimum of 21 (Kiowa)[17] and an average of 28.2. However, since there were only five languages of this type a statistical treatment would be of little value.

On the basis of the above information these types seem to be quite distinct. First consider the difference between the sets of languages having one and two component systems, that is one or two series of stops. The minimum number of consonants for the two types are: type I—nine, type II—thirteen. The maximum of consonants for the two types are: type I—28, type II—33.

[9] C. F. Voegelin, Inductively arrived-at models for cross-genetic comparisons of American Indian languages, UCPL 10.45 (1954).

[10] A. M. Halpern, Yuma I, Phonemics, IJAL 12.25–33 (1946).

[11] Paul L. Garvin, Wichita I, Phonemics, IJAL 16.179–84 (1950).

[12] Morris Swadesh, Salish Phonologic Geography, Language. 29.232–48 (1952).

[13] John R. Swanton, Linguistic Materials from Southern Texas and Northern Mexico, BBAE.127 (1940).

[14] Edward Sapir, The Phonetics of Haida, IJAL 2.143–158 (1923).

[15] H. J. Uldall, Maidu Phonetics, IJAL 19.8–16 (1953).

[16] J. Alden Mason, The Language of the Salinan Indians, University of California Publications in American Archeology and Ethnology, 14.1–154 (1918).

[17] William Wonderly, Lorna Gibson, and Paul L. Kirk, Number in Kiowa: Nouns, Demonstratives, and Adjectives, IJAL 19.1–7 (1953).

Finally the averages for the two types are: type I—14.5, type II—20.7.

These figures yield a difference in minima of four, in maxima of five, and in average of 6.2. Since it requires only two additional consonants (e.g. if system A has /p/, /t/, and /k/, then the addition of /b/ and /d/ would form a second series) to move system A from type I into type II, the average difference, if no factors other than chance were operative, would be expected to be somewhere near this figure. However, there is an average difference of slightly over six, which is an increase of more than 43 percent more consonants per language in type II than type I. It is extremely difficult to say just what would be statistically significant with materials of this type, but one would hardly expect to get an increase of 43 percent on the basis of sample selection.

Actually there is no reason for either the minimum, maximum, or average number of consonants to correlate with the a priori classification, which means then that four items, i.e. number of stop series, maximum number of consonants in any one structure, minimum number of consonants in any one structure, and the average number of consonants per language for the types, correlate much more highly than would be expected on the basis of a chance grouping.

The objection could be raised that systems having two series of stops would naturally have more consonants than would systems having one series. This may seem reasonable at first, but since it only takes four stops, properly selected, to form two series, and there are always other consonants besides stops, it is quite possible to envisage two component languages with only nine or ten phonemes in the consonantal systems. This makes the fact that two component languages have more than forty percent more consonants, on the average, than one component language, appear to be a significant item.

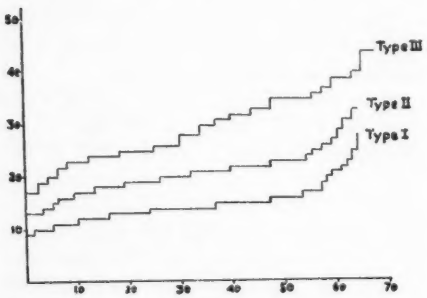
Consider now the differences between types II and III; the minima are: type II—thirteen, type III—seventeen; the maximums are; type II—33, type III—44; and the averages: type II—20.7, type III, 29.5. Again it only takes two stops properly selected to form a third series, thus moving a given language from type II into type III. However, the observed difference is 8.8 consonants per language more in type III than in type II. This is again an average increase of nearly 43 percent more consonants in type III than in type II. The fact that in both cases above, i.e. the difference between type I and type II, and the difference between type II and type III, the increase in average number of consonants per language is within five tenths of one percent of 43 percent is an unexpected parallel.

Since the differences between type I and type III would simply be the sum of the differences between I and II, and II and III, this will not be discussed in detail. It should be clear from the above discussion that type I and type III are quite distinct. Actually the average number of consonants in type III languages represents an increase of 98 percent over the average number of consonants in type I, this alone would seem to set the group off as distinct.

Graph IV below is a frequency distribution showing the number of consonants per language in increasing magnitude for each type. The lowest number of consonants in any system of type I is plotted in the lower left hand corner, then immediately above this is the lowest frequency in type II, and immediately above this is the lowest frequency in type III. Then going from left to right the next higher frequency for each type is plotted until the highest frequency for each type is reached at the upper right hand side of the graph. Since the number of languages in each type was not the same, the number of spaces covered by each language in the graph was adjusted so that the total distribution for each of the three types would cover approximately the same area.

The purpose of this plot is to demonstrate the fact that the above figures were

not deceiving. Statistics can obscure as well as clarify important facts about distributions, so this was plotted to show that there actually is a uniformly increasing number of phonemes, language for language, throughout each type, and not a heterogeneous sample, which when averaged together gave a misleading appearance of homogeneity. It would be quite possible to have different maxima, minima, and averages, yet have a single group of data with no empirical basis for sub-groupings. The averages could be made distinct by a few very divergent lan-



GRAPH IV

These histograms show the number of consonant phonemes for each system in each type starting with the minimum number on the left and reaching the maximum number on the right. Because of the differences in number of systems in the sample belonging to type III, each system covers twice as much area under the plot for this type as for types I and II.

guages, which when averaged in, gave the appearance of a higher or lower average, but this plot shows conclusively that this is not the case. There is a uniform, almost parallel, difference from the lowest point on the curves to the highest point. The shape of the lower histograms, which represent the number of consonants in each structure for types I and II, are almost identical. The upper histograms deviated only slightly from the general shape of the two lower ones, but this may be explainable in terms of the small number of languages of type III in the sample. The number of languages in the sample for types I and II is almost identical, while the number of languages in type III is only slightly over half that of types I and II.

Having shown by Graph IV the uniform differences between types I, II, and III, the next step is to examine statistically the frequency of occurrence of each consonant in each type. The purpose of this is to ascertain which phonologic features, i.e. the consonants, cluster in each type; or if the above correlation is the only feature by which the types can be differentiated. The reason for this is quite simple. By chance it is possible to find two unrelated items, in this case the average number of consonants in each group and the number of stop series, which correlate highly. However, the probability of finding three such unrelated features is considerably less, and with each added feature the probability of a chance correlation diminishes at a very rapid rate. What this means, in terms of typology, then, is that the correlations obtained are not simply a chance occurrence in this particular sample.

With each added correlation the probability increases that the types are distinct, each type having certain characteristics which set it apart from the other types, and no matter how much the sample is extended the types will still be valid. This does not preclude the finding of new types in other parts of the world, or adding features which are characteristic of the types demonstrated. It simply means that the probability of having these types prove to be unsound decreases considerably with each new correlation.

This investigation is not attempting, at this time, to theorize as to why these items occur together, but only to establish the fact that enough items in the phonologic systems of each type correlate to make it extremely improbable that this is a chance grouping based on the selection of the sample.

To recapitulate briefly what has been done, first the languages were grouped, a priori, on the basis of the number of stop

TABLE II

The frequencies of occurrence of the consonant phonemes in systems of type I listed in order of descending magnitude.

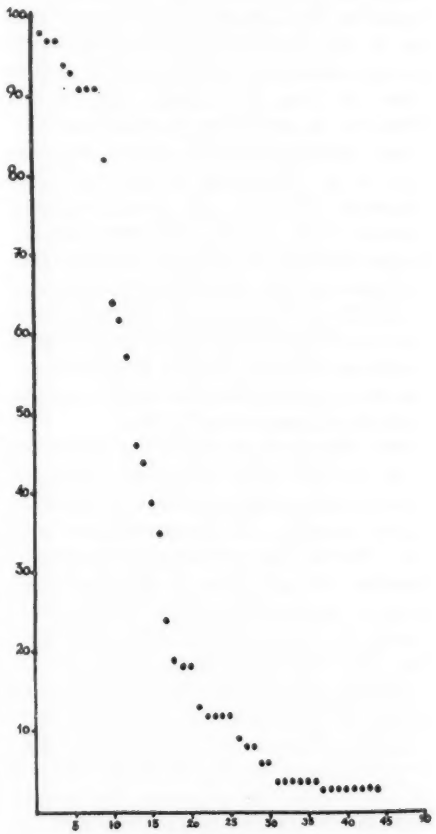
	Phoneme	Percentage of occurrence		Phoneme	Percentage of occurrence
1	n	98	23	ɬ	12
2	k	97	24	v	12
3	t	97	25	l^y	12
4	w	94	26	z	9
5	p	93	27	x̱	8
6	m	91	28	θ	8
7	y	91	29	ṯ	6
8	s	91	30	tɬ	6
9	h	82	31	d	4
10	š	64	32	x^w	4
11	tš	62	33	g	4
12	l	57	34	ḇ	4
13	ts	46	35	ð	4
14	ʔ	44	36	p	4
15	r	39	37	ž	3
16	x	35	38	$k̲^w$	3
17	n^y	24	39	r̄	3
18	ŋ	19	40	R̄	3
19	k^w	18	41	n'^y	3
20	k̲	18	42	ŋ̣	3
21	f	13	43	t̲	3
22	b	12	44	t^y	3

series in the consonant system. Next it was shown that the maximum, minimum, and average number of consonants differed considerably from type to type. Finally, lest it should be suspected that the statistics were introducing an apparent difference while obscuring real similarities, graph IV was constructed to show that each type was represented by a group of languages which formed a distribution, essentially parallel and similar in kind, but with type II differing—language for language—throughout the distribution by approximately forty percent from type I, and type III differing—language for language—by approximately forty percent from type II.

To add strength to the hypothesis that this grouping collects into each type consonantal systems which tend to share certain phonologic features, the systems of each group will be analyzed to see if certain consonants are typical of type I languages, while certain others are typical of types II and III.

To test the hypothesis that, type I languages tended to have certain consonants which occurred with less frequency in types II and III, Table II was constructed giving the percentage of occurrence for each consonant found in two or more type I systems.

To better illustrate the clustering of the points in the plot and the shape of the distribution, Graph V was constructed by plot-



GRAPH V

This plot represents the frequencies of the consonants occurring in systems of type I, plotted in order of descending frequency of occurrence.

ting the number of consonants in order of descending magnitude against their percentage of occurrence on cartesian coordinates. Graph V is similar to Graph I, but represents the frequency of occurrence of the consonants in type I only, not the entire sample. This distribution has certain characteristics in common with the one in Graph I.

First there is the cluster of /n/, /k/, /t/, /w/, /p/, /m/, /y/ and /s/ at the top of the graph with percentages of occurrence of 98, 97, 97, 94, 93, 91, 91 and 91 respectively. These are the same phonemes that are typical of all languages in the sample, with loss of /h/, which in dropping down to 82 percent occurrence in this type, is farther below /s/ than /š/ is below /h/, i.e. the difference in percentage of occurrence between these phonemes is greater than the sum of the differences for the first eight phonemes. However, /š/ is still eighteen percent below /h/, i.e. over twice the difference between /s/ and /h/, and two percent greater than the difference from /n/ to /h/ (the entire drop over the typical consonants). Thus the difference in percentage of occurrence between the two groups of consonants is greater than the range over the entire group, even counting /h/.

Now dropping down to /nʸ/ which occurs in 24 percent of the sample for this type, there is again a tight clustering as the curve swings smoothly down toward zero as a limit. The drop in percentage of occurrence between /x/ and /nʸ/ is eleven percent, whereas the maximum drop between successive consonants below /nʸ/ is five percent, just under half of eleven.

Having taken away the first and third groups of consonants, this leaves /š/, /ts/, /l/, /tš/, /ʔ/, /r/ and /x/ with percentgaes of occurrence of 64, 62, 57, 46, 44, 39, and 35 respectively. The consonant /h/ belongs to the upper cluster, in view of the eighteen percent drop to /š/, but this is not as clear cut as for the sample as a whole.

Here again there seems to be a group of typical consonants, i.e. /n/, /k/, /t/, /s/, /w/, /y/, /m/, /p/, and /h/, a rather long list of rare consonants, which begins at 24 percentage of occurrence and swings gradually off in a smooth curve approaching the 'x' axis as an asymptote. Then the seven consonants /š/, /ts/, /l/, /tš/, /ʔ/, /r/, and /x/ are scattered in an uneven distribution in the middle.

The first and third groups described above clearly set themselves off from the middle group; however, there are two possibilities as to where the break separating the typical from the non-typical consonants could be made, i.e. between /h/ and /š/ or between /x/ and /nʸ/. The former possibility has the disadvantage that the consonants typical of type I will be the same as those typical of the entire sample. A linear curve can be fitted rather well to the first group of consonants, and a logarithmic curve fits the last group, but the middle group seem to be almost randomly scattered. They conceivably could be an extension of the curve at the bottom of the graph, but this is not clear-cut.

In any event there can be no doubt that the first group of consonants, i.e. /n/, /k/, /t/, /w/, /p/, /m/, /y/, and /s/ form a tightly knit cluster of elements which are different from the rest. The phoneme /h/ is of course questionable, but because of the much greater drop between /h/ and /š/ than between /s/ and /h/, it is included with the first group. These then are the typical consonants for type I. It is unfortunate that these are the same as those for the entire sample, but by considering only these typical, none of the typical consonants occur in less than four-fifths of the languages of this type.

Chart III below then is a statistical model for the consonant systems of type I languages.

The next step was to examine the two component systems, i.e. type II. This examination of the correlation chart revealed that there appeared to be a slightly different group of consonants correlating with languages which have as their two com-

(p)	(t)	(k)
	(s)	
(m)	(n)	
(w)	(y)	(h)

CHART III

The consonants which are typical of the sample as a whole are inclosed in parenthesis. In this case that includes all of the typical consonants.

TABLE III

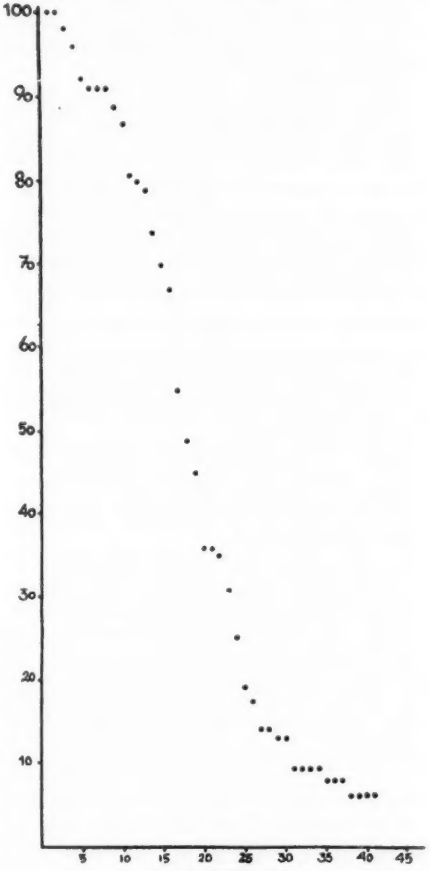
This table contains the phonemes which occur in type IIA systems with a frequency of six percent or more, which represents an occurrence in three or more systems of this type.

Phoneme	Percentage of occurrence	Phoneme	Percentage of occurrence
k	100	ŋ	35
p	100	f	31
t	98	v	25
n	96	θ	19
s	92	ž	17
m	91	n^y	14
b	91	ɬ	14
g	91	dz	13
h	89	ţ	13
d	87	x̣	9
y	81	l^y	9
w	80	g̣	9
tš	79	ṣ	9
š	74	k^w	8
l	70	g^w	8
r	66	d^y	8
x	55	tɬ	6
ts	49	g'	6
dž	45	ḍ	6
ʔ	36	t^y	6
z	36		

ponents glottalized vs non-glottalized stops, from the majority of type II languages, which have as their two components voiced vs non-voiced. Because of this apparent difference, two sub-types of type II were examined.

Table III gives the percentages of occurrence for all of the phonemes represented in three or more phonemic systems of type IIA, i.e. those systems having a series of voiced stops contrasting with a non-voiced series. It can be seen from the table that type IIA contains sixteen typical consonants: /k/, /p/, /t/, /n/, /s/, /m/, /b/, /g/, /h/, /d/, /y/, /w/, /tš/, /š/, /l/, and /r/ with percentages of occurrence of: 100, 100, 98, 96, 92, 91, 91, 91, 89, 87, 81, 80, 79, 74, 70 and 66 respectively.

Graph VI is a plot of the distribution, similar to those for the sample as a whole and type I in Graphs I and V respectively. This distribution has a relatively tight clustering of the above listed sixteen consonants i.e. a maximum drop between successive



GRAPH VI

The plotted points represent the percentages of occurrence of the phonemes found in type IIA systems in order of descending magnitude.

consonants of six percent. This is approximately one half of the eleven percent, drop between /r/, with 66, and /x/, with 55 percentages of occurrence.

The plot below the break has a maximum drop between consonants of nine percent and seems to have a different shape than that above. Because of the break in the distribution between /dž/ and /ʔ/ it is possible to make a case for adding /x/, /ts/, and /dž/ to the consonants typical of type IIA. However, in view of the fact that of these three only /x/ has a percentage of occurrence above fifty percent, they are not so considered here.

If the curves were fitted to the distributions there seems to be a slight curvature away from a single continuous curve, but this is not sharp enough to be conclusive. However, it seems that the break near the center, plus the change in curvature of the two plots near this break indicate that the consonants above this point could legitimately be set off as typical of type IIA. All the typical consonants in this type occur in over two-thirds of the languages in the samples, so it is quite proper to say that they are typical of the group.

Chart IV below contains the typical consonants for this type, again with the consonants typical of the sample as a whole enclosed in parentheses.

In comparing the typical consonants of type I and type IIA, both typical systems contain: /p/, /t/, /k/, /s/, /m/, /n/, /w/, /y/, and /h/. The consonants /b/, /d/, and /g/ occur in type IIA, as well as /tš/, /r/, /š/, and /l/. Thus there are three consonants which differentiate these two types which mean little in terms of correlation, as they

b	d	g
(p)	(t) tš	(k)
	r (s) š l	
(m)	(n)	
(w)	(y)	(h)

CHART IV

Consonants typical of type IIA.

TABLE IV

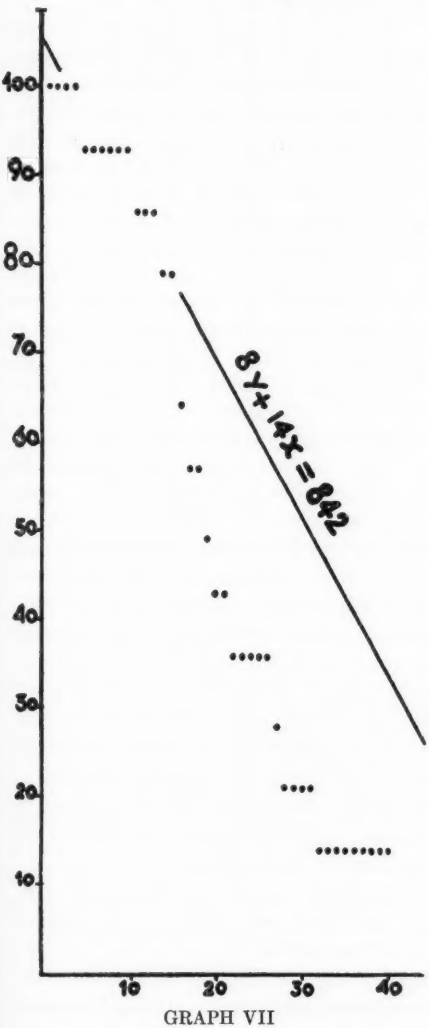
This table contains the phonemes found in type IIB, i.e. systems having a vocieless series of stops contrasting with a glottalized series, and their frequency of occurrence. Since a single occurrence represents a percentage of seven percent, nothing below fourteen percent was recorded.

Phoneme	Percentage of occurrence	Phoneme	Percentage of occurrence
n	100	t'š	43
k	100	r	36
t	100	k^w	36
t'	100	ķ	36
y	93	ķ'	36
p	93	x^y	36
w	93	n^y	28
h	93	b	21
s	93	ŋ	21
k'	93	tɬ	21
m	86	k'^w	21
l	86	v	14
p'	86	t'ɬ	14
ts	79	θ	14
ʔ	79	k'	14
tš	64	ţ	14
š	57	l'	14
t's	57	ķ'	14
ɬ	49	x^w	14
x	43	r̃	14

are diagnostic of the type IIA, i.e. /b/, /d/, and /g/. However, the other consonants are no more expected to correlate with this type than with any other. The result is then that there are four consonants:/tš/, /r/, /š/, and /l/, which correlate much more highly with systems having two series of stops than would be expected.

Consider now type IIB, which is characterized by a series of glottalized consonants, contrasting with a non-glottalized series. There were only fourteen languages of this type in the sample. Table IV contains all of the consonant phonemes occurring in more than two languages, or fourteen percent, of this type. This is a very small sample, partially the result of subdividing type II. It is so small in fact that this portion of the results are very uncertain indeed, but rather than obscure the picture, the results are given here with the admonition that the sample is small for significant results.

Graph VII is a plot showing the distribution of the percentages of occurrence of the consonants for languages of type IIB. Examining the plotted points, note the stair-step effect caused by groups of points all plotted at the same percentage of occurrence. This is the result of the smallness of the sample. If more languages had been available with this type of stop system the plot would smooth out into a regular curve. The reason is quite simple, a difference in frequency of occurrence of one (1) makes a percentage difference of seven (7).



GRAPH VII

The frequencies of occurrence of the phonemes in type IIB plotted in order of descending magnitude.

The break in the distribution between /ʔ/ with a percentage of occurrence of 79, and /tš/, with a percentage of occurrence of 64 is fifteen percent. Above this break the plot is linear, and the curve $8y + 14x = 842$ was fitted to the plot mechanically. This merely means that the slope of the line was ascertained by taking two points on the distribution through which a line could be drawn which gave a relatively close fit, i.e. points (3, 100) and (11, 86) and taking the two differences, $(11 - 3)$ and $(100 - 86)$. The slope then is $\frac{100 - 86}{11-3}$ or $\frac{14}{8}$. Now by substituting this into the general formula for a straight line, i.e. y = (slope) $x + a$ (where a equals 'y' when 'x' equals zero), the equation $y = -\frac{14}{8}x + 105.2$. Then by multiplying through the equation by 8, and transposing so that all quantities are positive, the equation $8y + 14x = 842$ (actually 841.6, but rounded off to the nearest whole number it gives 842).

Considering now the first fifteen consonants, i.e. those above the break, in order of descending magnitude of percentage of occurrence: /n/, /k/, /t/, /t'/, /y/, /p/, /w/, /h/, /s/, /k'/, /m/, /l/, /p'/, /ts/, and /ʔ/, these have percentages of occurrence of: 100, 100, 100, 100, 93, 93, 93, 93, 93, 93, 86, 86, 86, 79, and 79 respectively. These points fit a linear function very well. The linear function above, i.e. $8y + 14x = 842$ was fitted mechanically to the distribution. By extending this curve it is easily seen that the plotted points which represent the percentages of occurrence of consonants /tš/ and below, i.e. percentage of occurrence of 64 or below, do not even approximately fit the curve.

Because of the drop in percentage of occurrence, that is the fifteen percent break

p'	t'		k'	
(p)	(t)	ts	(k)	?
		(s) l		
(m)	(n)			
(w)		(y)		(h)

CHART V

Consonants typical of type IIB systems.

between /?/ and /tš/, and the fact that above this point the curve is linear, while being logarithmic below, these two segments were considered as two different distributions. Those above and including /?/ were set up as typical for type IIB. Chart V below contains the typical system of consonants for this type.

As for the preceding charts, the consonants enclosed in parentheses are typical of all languages. In addition to these three are: /p'/, /t'/, and /k'/ and possibly /?/, which are not typical of any types I or type IIA, but which could be considered as diagnostic for type IIB. Thus the correlation of these elements with this type is of no particular significance, except that it is interesting that so far as two series languages in the sample were found they were either voiced vs non-voiced, or glottalized vs non-glottalized. It is an interesting fact about languages of the new world that when they have two series of stops the most frequent contrast is between voiced and non-voiced and that the second most frequent is between glottalized and non-glottalized. It is worth mentioning that this exhausts all of the two component languages found in the sample. Maidu is the one possible exception, if implosion is considered the contrasting feature rather than voice, though the implosive series is also voiced.

In addition to the above differences, type IIB is distinguished from type I in that the typical system contains the consonants /ts/, and /l/. It is distinguished from type IIA by the fact that the typical system for the latter type does not contain the consonants: /p'/, /t'/, /k'/, /?/, and /ts/. In addition to this the typical consonant system of type IIA contains /b/, /d/, /g/, /tš/, /r/, and /š/, which are not found in the typical system for type IIB.

In addition to the diagnostic features, the typical consonant system for type IIA differs from the typical consonant system for type IIB by five consonants, i.e. /ts/, /tš/, /r/, /š/ and /?/.

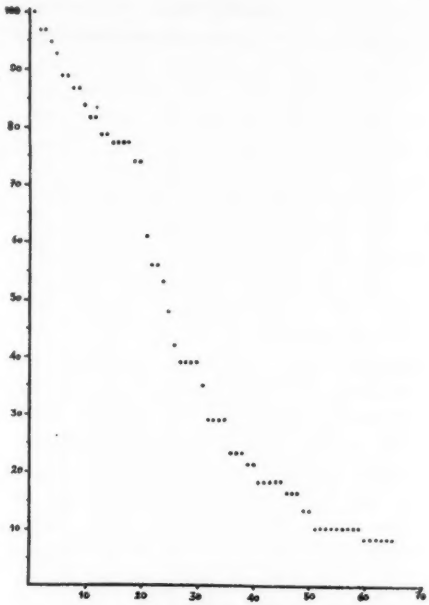
The above evidence seems to be quite enough to set these two groups of languages up as separate subtypes. However, there is an interesting question which arises from the typical structures of types I, IIA, and IIB. The question is should type IIA be classified as a subtype of type II, or should it be set up as a separate type? If the a priori grouping is followed it must be a sub-type of type II. However, the structure (typical consonantal structure as defined by the above procedure) indicates the possibility that it could be set up as an independent type. This implies of course that these classes, while observable groups which have distinct characteristics in a statistical sense, may be in reality based on something more subtle than the number of series of stops, but something that correlates highly enough with the stop systems that these can be treated as if they were the determining factor. However, these languages fit into type II on the basis of gross number of consonant phonemes per language, as well as number of stop series, so the class is labelled type IIB and considered to be a sub-group of type II. The possibility of establishing all three of the above groups as independent types has not been investigated thoroughly for a number of reasons, the most important of which is the smallness of the sample.

Going now to type III, the sample consists of 38 languages. There was no apparent subdivision of this type, and the group was treated as a unit. Table V contains all of the consonants occurring in type III systems with a percentage of occurrence above eight, i.e. more than two systems in the type. Graph VIII is a plot showing the distribu-

TABLE V

This table contains all of the phonemes occurring in Type III languages with a frequency of three or more or a percentage of occurrence of eight or more.

Phoneme	Percentage of occurrence	Phoneme	Percentage of occurrence	Phoneme	Percentage of occurrence
t	100	ɬ	56	g^w	18
k	97	dž	53	p'	16
s	97	t'š	48	z	16
n	95	ķ	42	ţ	16
h	93	r	39	t'š	13
m	89	ķ'	39	m'	13
y	89	dz	39	θ	10
w	87	t'ɬ	39	g	10
l	87	tɬ	35	ķ'	10
d	84	ŋ	29	ǧ	10
k'	82	k^w	29	x̣	10
t'	82	x̣	29	$x̣^w$	10
p	79	x^w	29	ķ'	10
tš	79	z	23	$ķ^w$	10
š	77	t'	23	$ķ'^w$	8
x	77	k'^w	23	v	8
ts	77	dl	21	l'	8
g	77	g	21	w'	8
ʔ	74	n^y	18	N	8
b	74	k'	18	ɬ'	8
p'	61	l^y	18	g^w	8
t's	56	n'			



GRAPH VIII

This distribution represents the frequencies of occurrence of the consonants of type III plotted in order of descending magnitude.

tion of the percentages of occurrence taken from Table V.

In Graph VIII there is a very sharp break between the 74 percentage of occurrence for /b/ and /ʔ/, and the 61 percentage of occurrence for /p/. This break is clearer actually than for any other group. Besides this sharp gap in the distribution, both the direction of the curvature and the general shape of the plots are quite different. If the distributions were extended beyond their present limits the extensions would miss the other plot considerably, both in shape and distance from the curve to the plotted points.

The consonants which form the upper distribution are: /t/, /k/, /s/, /n/, /h/, /m/, /y/, /w/, /l/, /d/, /k'/, /t'/, /ts/, /š/, /x/, /tš/, /g/, /p/, /ʔ/, and /b/ with percentages of occurrence of 100, 97, 97, 95, 93, 89, 89, 87, 87, 84, 82, 79, 79, 77, 77, 77, 77, 74, and 74 respectively. These then are the typical consonants for type III. Chart VI below, contains the typical system for type III, which is composed of these consonants. Again the consonants which are typical of the sample as a whole are enclosed in parentheses.

b	d			g	
	t'			k'	
(p)	(t)	ts	tš	(k)	ʔ
		(s)	š l	x	
(m)	(n)				
(w)			(y)		(h)

CHART VI

Typical consonants for type III systems.

Principally this type differs from all other types in having three series of stops; voiced, glottalized, and non-voiced-non-glottalized. However, note the absence of /p'/, which has a frequency of occurrence of 61 percent. This point is brought out only to indicate

TABLE VI

Contrastive list of typical consonants for all types.

Type I	Type IIA	Type IIB	Type III
p	p	p	p
t	t	t	t
		ts	ts
	tš		tš
k	k	k	k
		ʔ	ʔ
	b		b
	d		d
	g		g
		p'	
		t'	t'
		k'	k'
	r		
s	s	s	s
	š		š
	l	l	l
			x
m	m	m	m
n	n	n	n
w	w	w	w
y	y	y	y
h	h	h	h

that the fact that a series of glottalized stops, which is typical of a group or type, does not imply the occurrence of any specific glottalized stop. This means that the high correlation between glottalized stops and other consonants in type IIB above, even though occurring in a glottalized series, is more significant than might at first appear. For instance in order to have a glottal series here, it is necessary only to have two glottalized stops, but the fact that /t'/ and /k'/, with percentages of occurrence of 82, occur in a much higher percentage of the languages in this type than /p'/, with a percentage of occurrence of only 61, serves to demonstrate that the systems of typical consonants are not pre-determined by the method.

As stated above, the method implies that there will be three series of stops, in each language of type III, but it does not imply that any particular three series will be typical of the type. Actually it is unexpected that glottalized and voiced stops occur with a frequency high enough to bring them into the typical distribution, because this distribution requires a percentage of occurrence above 74 for this type, and on a purely chance basis, each series, i.e. voiced, glottalized, aspirated, nasal, and non-voiced-non-glottalized-non-aspirated-non-nasal, would occur in only 60 per cent of the languages of the type, because there are five series of stops which occur, and if three series are found in each language, this would mean that each series would appear in three-fifths, or 60 percent, of the sample for this type.

Even a cursory survey of the data demonstrates clearly that these series do not occur as would be expected by chance, but that certain consonants occur with a remarkably high frequency, while others are indeed quite rare.

Returning once more to the features which differentiate this type from the three preceding ones, type III contains only /x/ which does not occur in one of the above types. However, type III differs from type I in containing: /b/, /d/, /g/, /t'/, /k'/, /ts/, /tš/, /ʔ/, /š/, /l/, and /x/; from type IIA by containing /t'/, /k'/, /ts/, /ʔ/, and /x/ and by lacking /r/; and from type IIB by containing /b/, /d/, /g/, /tš/, /š/, and /x/ and by lacking /p'/.

In order to show more clearly the difference from type to type, Table VI was constructed. This table contains in contrasting lists the consonants typical of each type.

These differences cannot possibly be said to be a result of the method, because it would be equally possible, for example, using the above described method to have had /mb/, /nd/, /ndž/, /ng/, /p'/, /t'/, /k'/, /t's/, /t'š/, /t's/, /t'š/, /dz/, /dž/, and /p'/, as typical consonants for any or all of the types in the sample. There is also no reason that is at all obvious as to why the method could cause /r/ to correlate with type IIA and not with the other types, or why /x/ should correlate with type III and not with the other types. Finally there

is no reason why any one of the group shouldn't have had an aspirated series which contrasts with an unaspirated series rather than the predominant voiced and second most frequent glottalized contrast which occurs. Nor is there any reason why the consonants /x/ and /ʔ/ should correlate highly with type III and not with IIA, or /ʔ/ correlate to a high degree with type IIB while /x/ does not.

The only conclusion which seems tenable then is that these particular phonemes correlate considerably more highly than would be expected by chance. In order to test the method thoroughly a larger selection of languages will have to be made, and the same procedure used on this new sample.

However, regardless of how it would work on a larger sample, it has demonstrated some features which are typical of North American Indian languages as a group, besides indicating that the a priori grouping based on number of series of stops in the consonant systems of languages actually has a cluster of traits which correlate very highly with each specific group. Thus the arbitrary classification is not as arbitrary as it at first appeared, at least not so for American Indian languages.

Finally type IV contains phonemic systems having four series of stops. The sample of this group is so small as to make a statistical treatment impossible. There are only five languages of this type; i.e. Taos, Kiowa, Yana, Salinan, and Chirichaua Apache. The languages of this type do not appear to have a great deal in common with type III. However, no discussion, other than to say that this type seems to have more in common with type I than any other type, will be made.

2.4. Next consider Table VII, which contains the frequencies of occurrence of a small number of selected consonants in each type, as well as in the entire sample. These specific consonants were selected because the differences in frequencies of occurrence between types seem to give added weight to the typology. Small variations from type to type would be expected, but the differences in frequencies seemed unusually great in some instances, e.g. /tš/ occurs in two-thirds of type IIB and over three-fourths of type III, but in less than half of types I and IIA. Also /x/ occurs in one-third of type I, approximately one-half of type IIA and IIB, but in three-fourths of type III. Of course it is to be expected that the glottalized and voiced stops occur in type III with a higher frequency than in other types.

It is very difficult to evaluate this data, but there are no single series languages which have a single series of glottalized stops, in place of the typical non-voiced-non-glottalized-non-aspirated series. Many reasons why a language would not be likely to have a glottalized series as the only series, when no contrasting series occurs, could probably be set forth, but regardless of the reason, when a language has one stop series it is simply unaspirated and voiceless. If a language has two series, they tend to be a voiced series contrasting with a non-voiced one (72 percent of the two component systems have this type of contrast). Why is this the case? Why is there not a greater percentage of glottalized vs non-glottalized and aspirated vs non-aspirated? No answers have been proposed for these questions, nor do any appear in the data. The above treatment of these languages serves merely to bring the questions into sharp focus.

Looking again at the table, note the frequencies of the affricate /ʦ/, for type I of 46 percent, IIA of 49 percent, IIB of 79 percent, and III of 77 percent. This difference of approximately thirty percent indicates that this consonant correlates much more highly with types IIB and III than IIA and I. Consider further the affricate /tš/, which has a percentage of occurrence for type I of 62 percent, IIA of 79 percent, IIB of 64 percent, and III of 79 percent. This consonant correlates very highly with IIA and III, and not so highly

with I and IIB. The glottal stop /ʔ/ has a percentage of occurrence for type I of 44 percent, IIA of 36 percent, IIB of 79 percent, and III of 74 percent.

The above consonants are just a few examples to illustrate the variation found between types. The chart above permits such generalizations as this: /b/, /g/, /ts/, /tš/, /ʔ/, /š/, and /x/ occur in approximately three-fourths of type III languages. This means then that for any language having three series of stops, the chances are about three to one that it will also contain these particular consonants.

Another interesting item which shows up from an examination of the chart is the fact that type IIB languages contain the bilabial stop /b/ in about one language in four, whereas /d/ is not found at all in this type, and /g/ is found in only one language in ten. This means that in languages of type IIB, if a single voiced stop occurs, it is a bilabial one to a much higher degree than the velar one, and the dental stop is entirely absent. In contrast with this, in type III, which contains a voiced series in 84 percent of the sample, the voiced dental stop /d/ occurs in every language having a voiced series, except one—Bear River dialect of Athapascan.[18]

To continue the discussion of the table /t'š/ occurs in 48 percent of type III and 43 percent of type IIB languages, whereas it is almost totally absent from type I and type IIA. The phoneme /t'ɬ/ occurs in 39 percent of type III languages, but only about one-fourth as frequently (ten percent) in type IIB systems, and only once in all of type I and IIA. Since type IIB is characterized by a glottalized series the above consonants would be expected to occur here with a much higher frequency.

To conclude the discussion of Table VII, consider probably the most striking correlation of all. This is worthy of note because since these consonants /l/, /ḳ/, and /xʷ/ are all non-voiced, non-aspirated, and non-glottalized, and languages of all types are characterized by this group, these would certainly not be expected to correlate more highly with one type than with another. As a matter of fact, since three component systems by virtue of having three series of stops, normally have a greater number of consonants than do systems of types I, IIA, and IIB, one might even expect a slightly higher frequency of occurrence in the latter types in order to give the speakers of the languages in question more possible contrasts. However, the facts are just the reverse of this. These three consonants (i.e. /ɬ/, /ḳ/, and /xʷ/), while not occurring with a high enough frequency to be considered typical of the type, do occur with a relatively high frequency in type IIB and type III. The voiceless lateral /ɬ/ occurs in 56 percent of type IIB, but only 14 percent of type IIA, and zero percent of type I. The percentages of occurrence for /ḳ/ and /xʷ/ are around 40 percent for the former in type IIB and type III, but near zero for type I and type IIA, and the latter is 14 percent for type IIB and 29 percent for

[18] Pliny Earle Goddard, The Bear River Dialect of Athapascan, University of California Publications in American Archeology and Ethnology, 24.291–324 (1929).

TABLE VII

The figures below show the percentages of occurrence for each of the phonemes listed in types I, IIA, IIB, III, and the sample as a whole. The purpose is to illustrate the relative frequency differences from type to type.

	p	b	d	g	p'	t'	k'	ts	tš	ʔ	š
Type I	93	12	04	02	00	00	00	46	62	44	64
Type IIA	100	91	87	91	00	00	02	49	79	37	77
Type IIB	93	21	00	10	86	100	93	79	64	79	57
Type III	79	74	84	77	61	82	92	77	79	74	77
Sample	92	55	51	49	25	29	29	58	71	52	64

	l	x	t's	r	ɬ	t'š	k	dz	t'ɬ	xʷ
Type I	57	35	00	39	12	00	00	00	02	04
Type IIA	70	58	04	66	14	02	04	13	00	00
Type IIB	86	43	57	36	49	43	36	05	10	14
Type III	87	77	56	39	56	48	42	39	39	29
Sample	71	53	22	49	26	20	21	14	09	08

type III, but near zero for type I and type IIA.

The above discussion of the relative percentages of occurrence in different types was simply to illustrate that there are quite a large number of phonemes which while not typical of any particular type still occur with a much greater frequency in one type or another than would be expected purely on the basis of chance.

For example, there is a heavy clustering of velar and palatal consonants including /ḳ'/, /g̊/, /x̊/, /gʷ/, /g̱/, /x̱ʷ/, /ḳ'/, /ḳʷ/, /g̊ʷ/, and /ḳ'ʷ/ in type III. These are almost absent in types I, IIA, and IIB, and do not show up in the typical charts, because they are very low frequency consonants, even in type III (varying from eight percent to 21 percent). However, they occur much less frequently in type I (/ḳʷ/ in only two systems of that type, and the remainder in one or no systems of type I). In type IIA systems /g̊ʷ/ has a percentage of occurrence of eight, and the remainder below two percent. In type IIB systems /ḳ'/ has a percentage of occurrence of fourteen, which actually represents an occurrence in only two systems of this type. Finally, /x̊/, /x̱ʷ/, and /ḳʷ/ with percentages of occurrence of seven, or one language, and the remainder do not occur at all in type IIB.

Certain other consonants among those included in Table VIII, while not appearing in the typical systems of any type do occur with an unexpectedly high frequency in type IIB and type III, as opposed to the type I and IIA. The remainder of the consonants in Table VIII appear in the typical systems of one or more of the types, and were included only to show the varying percentages of occurrence from type to type.

To conclude this discussion, a brief review is in order. This was a statistical study of the consonantal systems of 176 languages. The purpose of the investigation was to test the hypothesis that a phonological typology based on the number of stop series correlated sufficiently with the specific elements—i.e. the consonants themselves, of the systems to make such a typology meaningful. The first step in the procedure was to gather the data and reduce the variety of symbols to a standard set (cf. **2.1**), and evaluate the quality of the materials. The systems which were derived by the investigator can best be described as allophonic, rather than phonemic (cf. **1.2**). However, because of the number of languages investigated and the quality of the data, the same sample rigidly phonemic would probably give the same results. The languages of the sample were classified on the basis of the number of stop series in each language, and the frequencies of occurrence for each consonant in the sample as a whole were counted. This procedure resulted in the separation of the consonants into: typical, nine consonants, i.e. /k/, /t/, /n/, /s/, /m/, /p/, /w/, /y/, and /h/, which occur in virtually every language of the sample; rare, approximately one hundred consonants which occur in less than thirty percent of the languages; and an intermediate group /ts/, /l/, /š/, /tš/, /b/, /w/, /ʔ/, /d/, /g/, and /r/, various constellations of which form the systems typical of the various types, with all types containing the consonants which are typical of the sample as a whole. In general the intermediate group accounts for the consonants that are typical of the various types, but not the sample as a whole. The only consonants which serve to differentiate types, that are not included in this group are the glottalized /p'/, /t'/, and /k'/. It is difficult to explain why this is the case, but these three consonants have a percentage of occurrence of only; 25, 29, and 29, respectively. Possibly a larger sample would push these up into the intermediate group, but in this sample they definitely belong to the group of rare consonants.

To continue, since the above procedure

yielded such unexpected results, it was repeated on each of the above types, including sub-types IIA and IIB (cf. **2.3**). Again the percentages of occurrence were plotted for each type, and these distributions appeared to set off a certain group of consonants as distinct from the rest in that they occurred with a very high frequency in that particular type. The evidence which indicates that these high frequency consonants are different from those having a lower frequency is the shape of the curves which were fitted to the plotted points representing the percentages of occurrence, and the fact that in all but one case there is a fairly sharp break in the distribution, with the shape of the curves being different above and below this break. In type I there is a break between /h/ and /š/, but there is also a rather large break between /s/ and /h/ (half difference between /h/ and /š/ leaving doubt only as to whether /h/ is typical of this group or not. Because the break between /h/ and /s/ is only half the break between /h/ and /š/, and the fact that /h/ is typical of the entire sample, it was considered to be typical of type I. The rare consonants form a closely arranged distribution starting at 24 percent and dropping smoothly down toward zero as a limit. Thus for type I, as for the sample as a whole, there were three kinds of consonants; typical, rare, and an intermediate group.

An attempt was made to correlate some aspects of the consonant systems for each type. Graph IV demonstrates rather clearly that the types, language for language, differ as to gross number of consonants per language as well as maximum, minimum, and average number of consonants per language. This is a very interesting and quite unexpected correlation. Types IIA and IIB, show no obvious differences in gross number or maximum or minimum number of consonants. This is one of the reasons for considering them as sub-types rather than independent types.

Thus the gross number of consonants per language in each type as well as the maximum, and the minimum number, are distinctly different for each type, and there is a certain constellation of consonants which is typical of each type, in contrast to the others. Finally in addition to the typical systems, there are a few consonants which, while not being typical for any one type, occur much more frequently in one than another (cf. Table VII).

Thus the evidence has been presented above which seems to indicate that what was originally a completely arbitrary grouping, based on the number of stop series in the consonantal system of a language, is in fact a meaningful typology.

NOTES AND REVIEWS

Methodological Considerations Regarding a Statistical Approach to Typologies

Sol Saporta

The present remarks are aimed at restating some of the problems involved in any statistical analysis of the consonant systems of a large number of languages, such as the analysis proposed by Joe E. Pierce, A Statistical Study of the Consonants of New World Languages, IJAL 23.36–45,94–108. Although the data collected by Pierce are in themselves of interest, the main contribution of his presentation is his suggestion for a method to 'test statistically the validity of . . . a grouping . . . [of] phonological systems on the basis of the number of series of stop consonants.' Before his results may be accepted as valid, however, some clarification of basic concepts seems indicated. At C. F. Voegelin's suggestion three such concepts are briefly discussed below (**1.** Comparability of data; **2.** Method; **3.** Typologies and correlations) followed by speculations and suggestions for further investigation (**4.** Alternate methods).

1. Of both theoretical and practical interest is the problem of to what degree phonological systems or sub-systems may be equated. The point has been clearly made by Weinreich:[1] 'the phonemes . . . are defined within each language by oppositions to other phonemes . . . of that language. For example, /p/ in Russian . . . is defined . . . by its distinctive feature of non-palatality, . . . while the definition of /p/ in English . . . involves no such restriction. From the point of view of the languages, therefore R[ussian] /p/ and E[nglish] /p/ cannot be "the same" '. Weinreich indicates however, that while the phonemes of two languages are in theory not commensurable, 'in practice classificatory overlappings of physical sound . . . are of course extremely common.'[2] It is certainly true that bilinguals, for example, do make these interlingual identifications, which in most cases will be perfectly apparent to the linguist. But the identifications by bilinguals do not always involve one-to-one relationships. Thus, an /ɛ/ in language A may be identified with /ei/ in language B; a unit phoneme /č/ in one language may be identified with a cluster /tš/ in another, etc. And yet, for the purpose of comparing inventories, we would be obliged to state that the phoneme /č/ was present in one language but absent in the other. In any case, it is clear that before any comparison becomes valid, care must be taken that the phonological systems themselves are being compared, and not merely the symbols used to represent the systems.

The same problem is involved in deciding on what symbol is to be used in representing a phoneme. If two allophones are in complementary distribution, it is conventional, in a phonemic transcription, to use the symbol which represents the allophone with the least restricted distribution. Thus, if [m] occurs only before bilabial consonants and [n] occurs in all other environments, the phoneme will usually be symbolized by /n/. The convention is not always followed, however, as indicated by the Spanish series of voiced phonemes /b d g/. Each phoneme has both a stop and a spirant allophone. In general, the stop occurs only after pause and after a nasal; the spirant occurs elsewhere. One should according to the above, write these phonemes /b đ γ/. (There are two voiceless series, the stops /p t k/ and the fricatives /f θ x/.) The decision to use the symbols /b d g/ to represent the voiced

[1] Uriel Weinreich, *Languages in Contact*, New York 1953, p. 7.

[2] Ibid., p. 8.

series in phonemic transcriptions is perhaps dictated by orthographic convenience or some other non-structural considerations. The result is that Spanish is usually characterized as having two series of 'stops'.[3] To what extent such problems were present in Pierce's data is, of course, difficult to estimate.

2. In essence, Pierce's method consists in identifying those phonemes which are 'typical' of all languages (Chart II), and then identifying those phonemes which are 'typical' of languages with one series of stops, languages with two series of stops, and languages with three series of stops. Thus, /k/ /t/ /n/ /s/ /m/ /p/ /y/ /h/ /w/ are said to be typical of all languages in the New World because they occur in from 99% to 86% of the languages examined. The next most frequent phonemes are /č/[4] and /l/ which occur in 71%. This 'break' (between 86% and 71%) is used to separate one class of consonants from another. Pierce maintains that this break is not merely fortuitous and points out that in a more rigorous analysis of the data /č/ would 'undoubtedly be a cluster' and '/l/ is probably an allophone of /r/ in some instances'. This presumably would reduce the percentage of /č/ and /l/, and thus make the break even sharper. If it is true that an accurate phonemic analysis would have reduced the number of occurrences of /č/ and /l/, then it is equally true that such an analysis would have reduced the number of languages with /w/ and /y/; surely there are at least some languages, where they may be considered members of /u/ and /i/. This would have precisely the opposite effect, namely smoothing out the distribution so that the break would be less apparent. Certainly, speculations about what more accurate data might indicate are no substitute for the data themselves, but, without entering into an evaluation of the statistical techniques, one cannot help getting the feeling that Pierce has not demonstrated that the distribution of the data is not very close to what one might have expected by chance.

3. It is clear (Graph IV) from Pierce's data that when consonant systems are classified according to the number of series of stops, then the number of consonants per language will be greater for languages with two such series than for those with one series; languages with three series of stops have a larger number of consonants than languages with two such series. It is not clear, however, why this is such an unexpected correlation. Languages of Type I have an average of 14.5 consonants; languages of Type II 20.7. Pierce maintains that 'since it requires only two additional consonants to move a system . . . from type I into type II, the average difference . . . would be expected to be somewhere near this figure.'[5]

The difference may be explainable by one of two factors. First, by utilizing a second series of stops, e.g., a voiced series, a language has available a new distinctive feature which is by no means restricted to the stops. Consequently a language with two series of stops is more likely to have two series of fricatives, than is a language with one series of stops.[6] Thus, in the list of phonemes in languages of type IIA (Table III), one finds, besides /b d g/, other

[3] See, however, George L. Trager, The Phonemes of Castilian Spanish, TCLP 8.217–222 (1939).

[4] Without discussing the relative merits of the cluster solution vs. the unit phoneme solution, it is perhaps better practice to use a single symbol to indicate the latter; consequently /č/ has been used here for Pierce's /tš/, etc.

[5] Pierce's argument seems a little like saying that if one measured 100 people over six feet tall and found an average height of 6′2″, and then measured 100 people under six feet tall and found an average of 5′8″, that the six inch difference is unexpected, because if 'properly selected' it would require only two inches to move a person from one group to the other.

[6] There may of course be exceptions. Cherokee apparently has only one series of stops, but has a contrast between /s/ and /z/.

voiced phonemes such as /z v ž/, etc., which while not frequent enough to be 'typical' of this group, certainly serve to raise the average number of phonemes. Second, and perhaps not so obvious, is the implication that before a language introduces a second series of stops it utilizes at least three places of articulation. There are no examples of four stops arranged in 2 x 2 patterns such as $\begin{matrix} p & t \\ b & d \end{matrix}$ or $\begin{matrix} p & t \\ p' & t' \end{matrix}$, whereas 1 x 4 patterns such as /p t č k/ are common. Now, examples are not evidence, but the hypothesis that differentiation in the type of stop is correlated with differentiation in the place of articulation merits further investigation.

Wells[7] has pointed out that 'a classification is natural to the extent that each class is characterized by properties not implied a priori by its defining property.' Thus, one would expect the correlation between the gross number of consonants in a system and the number of series of stops; one would not expect the correlation, if there is one, between the number of series of stops and the number of places of articulation used in the stops.

4. It is obvious that there is a certain arbitrariness in selecting the number of series of stops as a basis for classification. One could just as well classify according to the number of nasals, or the number of liquids; the result would be a different classification. The possibility suggests itself, then, that one might reach a classification with no prior assumptions as to what the basis for the classification should be. Ideally, one would first determine which languages 'go together' and then ascertain, if possible, which phonemes or set of phonemes are characteristic of each group, instead of the reverse.[8]

If we assume that phonemic inventories can be compared on a 'same-or-different' basis, e.g., that English /p/ is the 'same' as Russian /p/, but 'different' from Russian /p^y/ then the resulting data can be analyzed by several statistical techniques.[9] The most simple, but not necessarily the most adequate, is ascertaining the percentage of agreement between any two phonemic inventories by using the formula:

$$\frac{2(C_1C_2)}{C_1+C_2}$$

where C_1 is the number of consonant phonemes in language 1, C_2 is the number of consonants in language 2, and C_1C_2 the number of consonants common to both,[10] this last value being multiplied by two to yield a percentage value. If language 1 has 22 consonants and language 2 has 18, and if they have 14 in common, then the percentage of agreement would be:

$$\frac{2(14)}{22+18} = 70\%$$

It can be seen that 100% agreement can occur only when both languages have the

[7] Rulon Wells, Archiving and Language Typology, IJAL 20.103 (1954).

[8] Pierce's classification is essentially based on inventory alone. In addition to differences in inventories, consonant systems differ in permissible clusters, frequencies, and utilization of distinctive features. A classification might (and, perhaps should) use any of these other dimensions, or combinations thereof.

[9] See, for example, David W. Reed and John L. Spicer, Correlation Methods of Comparing Idiolects in a Transition Area, *Language* 28.348–359 (1952); J. P. Guilford, *Psychometric Methods*, New York 1936.

[10] This formula differs from the one suggested by Milewski, cited by Dell H. Hymes, in his review of *Lingua Posnaniensis* in IJAL 22.285 (1956), where '(A − B) ÷ (A + B) = X, where A is the sum of the phonemes having equivalents in the two systems, B the sum of the phonemes without equivalents in the second language, and X the index of phonological resemblance between the two languages.' If I understand Milewski's formula, the result may vary from +1.0 (perfect identity) to −1.0 (no phonemic equivalences). The percentage of agreement may vary from 1.00 to .00.

TABLE 1. Percentages of Agreement

Languages	1	2	3	4	5	6	7	8	9	10	11	12	13
1													
2	75												
3	69	86											
4	60	58	62										
5	56	61	68	38									
6	51	67	68	51	60								
7	50	67	74	50	64	85							
8	64	81	82	71	55	70	70						
9	62	79	74	62	54	68	68	86					
10	77	73	69	62	63	59	58	71	69				
11	54	63	56	49	65	58	57	67	65	74			
12	53	71	67	53	52	76	76	68	67	61	60		
13	53	68	80	59	57	68	78	76	70	60	53	73	

TABLE 2. A Partial Classification Based on Percentages of Agreement

Languages	2	3	8	9	10	6	7	12	13
2									
3	86								
8	81	82							
9	79	74	86						
10	77	73	71	69					
6									
7						85			
12						76	76		
13						68	78	73	

same number of consonants and agree perfectly in their identity, e.g.,

$$\frac{2(22)}{22+22} = 100\%$$

As an illustration of what kind of results such a procedure would yield, the following eleven languages were chosen, more or less at random:[11] Cree (1), Yaqui (2), Payas (3), Cherokee (4), Ona (5), Seti (6), Coroado (7), Nambikware-Tarunde (8), Chatino (9), Maidu (10), and Taos (11). In addition, two Indo-European languages, Spanish (12) and English (13) were included. The percentages of agreement are shown in Table 1.

[11] An effort was made to include examples of each of the major types suggested by Pierce.

Obviously, these data themselves do not yield clear-cut types whereby any language can be unambiguously assigned to one and only one type, and to what extent the data can provide a meaningful basis for classification is a question which remains to be answered. Nevertheless, if one requires a certain minimum agreement, say, 68% between any two languages in a type, certain classes do emerge, as indicated in Table 2.[12] According to such a presentation, languages 2, 3, 8, 9 and 10 are to be classified together, although, according to Pierce's classification, 2 and 3 have one series of stops, 8 has two series, and 9 and 10 have three series of stops. An inspection of the inventories shows that, as far as the stops are concerned, all languages have voiceless /p t k/ ± /ʔ/; + voiced /b/ ± /d g/; ± a third series. The total number of consonants ranges from 14 to 19.

The above is not meant to imply anything about the relative merits of the two methods. On the contrary, the point is precisely that there are a large number of criteria, and that alternate criteria yield alternate classifications. Nevertheless, it is doubtful whether these alternate classifications are equally valid, and what is needed is some attempt at devising a test of the relative validity of cross-genetic classifications. In short, if, given information about language 1, we can make greater-than-chance predictions about language 2, then we have evidence that the basis for classification is a meaningful one; at least such correla-

[12] Clearly, the smaller the required percentage of agreement, the greater the number of languages which will be grouped together, and the fewer the types needed to exhaust all the languages in the sample. However, this increases the number of 'border-line' languages, i.e., languages which could be assigned to more than one type. Another possibility, involving an undesirable arbitrariness, is to select one language (for example, the one with the smallest inventory) and use it as a basis for classification. Thus, if, say, Cree were chosen, all languages with a high percentage of agreement with Cree would be classified as members of the Cree type.

tions would reduce the degree of arbitrariness which characterize many typologies.

INDIANA UNIVERSITY

SPANISH HAMACA AND ITS CONGENERS

DOUGLAS TAYLOR

In his book *The Arawak Language of Guiana* (Amsterdam, 1928; p. 235), the late C. H. de Goeje wrote:

> It does not seem at all impossible, that the first European who saw an Indian hammock and asked the name for it, got the answer 'to sleep', A.M. *imaka, and that the word hamaka, which is met in a few vocabularies of A.M. languages, is really the Creole word. The Arawaks, though they know the word hamaka, have also a word of their own for a hammock.

This word, variously spelt 'ukura', 'okura', 'okora', though clearly cognate with its Island Carib equivalents, ékora (XVIIth-century Dominica) and -ógora (XXth-century British Honduras), is quite unlike Spanish hamaca *hammock*, which was borrowed from Taino. On the other hand, while I do not agree with de Goeje's reconstruction of the proto-form, his assemblage of Arawakan words for *sleep* makes it evident that most of them are cognates, and highly probable that Taino hamaka, whatever its meaning, is related to them. Compare: uatümaka (Manao), atũka (Goajiro), itiemeka (probably /itemeka/; Saraveca), temeka (Marauha), timana (Ariti), timka (Araicu), uatsima (probably /uatima/; Yavitero), tsima (Baniva), himaka (Palicur), himeka (Guana), hamaka (Ipeca), demakaini (Guinau), domakari (Bare), domakale (Arua), dau (Wapisiana), adonka (Lokono), arųka (Dominican Island Carib, XVIIth C), arúmuka (Dominican Island Carib, XXth C), arúmuga (British Honduran Island Carib, XXth C), remöka (Cujisenajeri), manimaka (Adzaneni), nuimaka (Amarizana), nimata (Passe), pikiemato (/pikemato/ by metathesis from /pitemako/? Yucuna), wemakya (Cauixana), uymaka (Jumana), iemaka (Tariana), imaka (Carutana, Ipurina, Piapoco, Siuci), imakaua (Catapolitani), imoko (Baure, Mojo), imake (Mandauaca), amahe (Campa), imama (Arequena), maka (Ariti, besides timana), magha (Mariate), mka (Piro),—all meaning *sleep*.

Now in Island Carib all verbs are derived (primary or secondary) words, consisting of a base, which may or may not also occur as a free form of another class, together with verbalizing prefix and/or suffix. And if I am right in assuming this to have been true of the proto-language (in which all such bases presumably might occur also as free forms), I take it that -ka of the words for *sleep* in so many Arawakan languages is or was a verbalizing suffix, and that base and verb coincide in Baniva tsima *sleep*. Moreover, comparison of these with other cognates belonging to the same languages (e.g., Palicur aha:Lokono ada:Yavitero ata *tree, stick or wood*; Palicur tipa:Lokono, Mandauaca and Yavitero siba:Baniva, Carutana and Yucuna hipa *stone*; Palicur isiki:Mojo iciki:Lokono and Campa itika *excrement*; Palicur tiketi:Saraveca tik(i)ahi: Mapidian čikasi:Goajiro sikih:Lokono hikihi *fire*) makes it probable that the proto-language had at least two and possibly three apical stops, *d, *t, *th, and that the base of the verb meaning sleep contained the first of these. Tentatively, I should reconstruct this base as *-duima-, giving Lokono -don- (< *-doma-) and Island Carib -rų- ~ -rumu- (<*-ruma-).

This immediately calls to mind, semantic difficulties notwithstanding, the themes: Lokono -(o)doma *because of*, Goajiro -(u)-tuma *by* (*the agency of*), Island Carib -(u)-ruma *by* (*the agency of*), *because of*, *for the sake of*, and Lokono -(o)ma, Goajiro -(u)ma, Island Carib -(u)ma, all meaning *together with*; both sets combining with the person-marker prefixes to form what in the last-named language are morphological nouns. If, as seems possible, members of the first set contain those of the second set, they may also contain a morpheme in common with Lokono -doli *root* and -dora *rib*, Island

Carib úrule and urú(h)ule (which have the same meanings). Elsewhere we find Lokono -hodo- *bent or bowed down* and Island Carib uru- *sunken* ('*enfoncé*'). Merging of the concepts 'lower level' and 'determining principle, support' is common in many languages, as in the English verb to found and its derivatives; and it is at least possible that a morpheme having both these senses combined with another meaning *in company of or relation to* to form the bases of what have become modern Central American Island Carib arúmuga *to sleep*, burúma *because of thee* (or *by thy agency*), together with these words' equivalents in many Arawakan languages, including Taino hamaka, whence borrowed Spanish hamaca and our English hammock.

DOMINICA, BRITISH WEST INDIES

LANGUAGES AND GHOST-LANGUAGES OF THE WEST INDIES: A POSTSCRIPT

DOUGLAS TAYLOR

In a note recently published under the above title, I said (IJAL 22, p. 182) that what little was known of the languages of Trinidad appeared to be contained in de Laet's short word-lists of Arawak, Shebayo, and Jaoi (or Iaoi). Fortunately, I was mistaken; for no sooner had I dispatched that ms. than R. W. Thompson of the U.C.W.I. (Jamaica) very kindly sent me copies of other lists and data, relevant details of which are given below.

IN THE VOYAGE OF ROBERT DUDLEY . . . TO THE WEST INDIES, 1594–1595 (Ed. George F. Warner; Hakluyt Soc. 2nd series No. 3; London, 1899), Dudley lists on p. 78 of his narrative 'certaine wordes of the language of Trinidad which I obserued at my being there' (67 entries); and Capt. Wyatt, who took part in the same voyage, gives 27 words of 'Aroaca, *sermo Indianus*' on p. 65 of his own narrative, contained in the same volume. The words of both these lists (as well as those of one of the three recorded in 1598 by de Laet) evidently belong to the same language, which appears to be identical with that called Lokono or True Arawak, now spoken in British and Dutch Guiana.

On the other hand, A. Vazquez de Espinosa gives scraps of various languages in Chapter XVI of his COMPENDIUM (published in 1629); and on p. 95 we find: 'La nacion de los Indios de la isla Trinidad, llamada Nepuya, y los de la Provincia de Guayana, hablan casi una misma lengua', followed by ten short phrases and words apparently belonging to a Cariban language. He then goes on to 'La nacion Arauaca que habita cerca de la bocas del rio Orinoco', and gives (there and on p. 186) twenty-one words of their language, which appears to be Lokono of an only slightly (if at all) different variety from that recorded by Dudley and Wyatt in Trinidad.

Since not all of the above-mentioned sources may be easily accessible to the interested reader, this material (except for the Jaoi list, a copy of which I have not got) is reproduced below. Native forms are given in the spelling of their recorder, as are Vazquez' Spanish translations; but Dudley's English and Wyatt's and de Laet's Latin are rendered into modern English. Nor do I list all the items in their original order. Abbreviations employed are: D (Dudley), W (Wyatt), L (de Laet), V (Vazquez). Forms given in parentheses without other mention are modern Lokono; those preceded by IC are Island Carib of XVIIth-century Dominica.

Shebayo of Trinidad, L: heja *father*, hamma *mother*, wackewijrrij *our heads* (Goajiro ekiwi *head*), wackenoey *our ears*, noeyerri (*my?*) *eyes*, wassi (uasiri) *our noses*, darrymaily (*my?*) *mouth* (IK iuma, Uirina luma, *mouth*), wadacoely *our teeth* (Wapisiana dako, Atorai dak, *tooth/teeth*), watabaye *our legs*, wackehyrry *our feet* (Atorai kheti *foot/feet*), ataly (ada) *tree*, hoerapallii *bow*, hewerry *arrows*, kyrtzyrre (Mapidian

kirsu, Mehinacu kerzi) *moon*, wecoelije *sun*.

Arawak of Trinidad: D dabodda, W dabádoh, (dabada, IC núbara), *my nails* (*claws*); D dabarra, dabarah, W dabárroh, (dabara), *my hair*; D dacan, dacabbo, W dacabo, (dakabo, IC núkabu), *my hand*; D dacosi, W dacasi, (dakusi, IC náku), L wackosije (uakusi) *our eyes*; D dacurle (dakoro) *my knee*; D dacutti, L dackosye, (dakuti, IC nugúti), *my feet*; D daddano, L dadane, (dadana, IC núrna), *my leg*; D dadena (dadone, IC naróna) *my arm*; D dudica, W dadica, (dadiki), *my ears*, L wadycke (IC uaríkae) *our ears*; D dalacoak (*sic*, misprint?), W dalarócoh, L dalerocke, (daleroko), *my mouth*; D areheh (ari) *tooth*, W dary, L darii, (dari, IC nári), *my teeth*; D daria *my gums*; D dasereth *my eye* (? cf. dasere, IC nesére, *my ulcer*); W daciboh (dasiba, IC nisíbu) *my face*; D desire *my lips*, W dacirey (dasiri, IC nísiri) *my nose*, L wassyerii *our noses*; D dessie *my forehead*, W dacy (dasi) *my head*, L wassijehe *our heads*; D dill (daie, IC niníe) *my tongue*; D addoth, W adda, L hada, (ada), *tree/stick/wood*; D haddalle, W hadaley, L adaly, (hadali), *sun*; D cattie, L cattehel, (kati), *moon*; D hicket (hikihi) *fire*; D oronuie (oni- ~ uine) *water*; D bara, W bara, *water*, (cf. bara *sea*); D weeuah, W weeuah, *sky*, (cf. uiua *star*); D sibath (siba) *stone*.

Idem: D tauraroth, W taiourah, (taho), *rope*; D ureit, W uree, (iuri), *tobacco*; W casaca *clouds* (cf. kasaku *sky*); W basya *wind*; D tacarao, W tácorah, (IC takúlaua), *greenstone* (*pendant?*); D caulpiri, W colperey, *a white stone*; D harowa *a stone good for the headache*; D curaballa *a kind of stone*; D barrenaire *beads*; D arrara, W arara, *copper*, (cf. IC álala *dusky, burnished, glistening*); D calcouri, W calcurey, (karokori, IC kalúkuli), *gold*; D mointiman *iron or steel*; D hemachugh *cassava-bread*; D callit (kali) *cassava-bread*; D halete (haliti) *sweet potato*; D maurisse (marisi) *maize*; D carewoda *a sweet root*; D mentinie *a kind of tree*; D guttemock *a man*; D howa (hua) *monkey*; D walrowa *parrot*; D ulasso *tunny-fish*; D bohery *flying-fish* (cf. buhiri, IC búliri, *bat*); D tibetebe *cockles*; D carotta *a thing like a pappe*.

Idem: D semaro, L symare, (simara), *arrows*; D marrahabo, L semarape, (simara(ha)bo), *bow*; D haculle *bow-string* (cf. uhikili *pita or silk-grass*, from which bow-strings were made); D yeddola, W eduóla, (iadoala), *knife*; D casparo (kaspara; cf. Sp. espada) *sword*; D techir *bracelet*; D barudda (balida, IC buléra) *comb*; D beyou (beiuka-hu) *flute*; D tolletillero *bells*; D queca (kiueke) *kind of basket*; D bodad (borodi) *box*; D sambolers *sic*, 's' for 'o'? sambuleru, from Sp.) *hat*; D arguecona, arkeano, W arkekano, (arukuana-hu), *scissors*; D addehegaeno (adekekona, adokakuana-hu) *spy-glass*; D heldaro (halalu, hararo) *pot-spoon*; D perota *silver* (Sp. plata); L pilplii *father*; L saeckee *mother*; D sakel (sakele < sa *good or well* + -kele, a generalizing suffix) *it is well or I am well*; D non quo, non quapa '*I know not, or I cannot tell*'; D 'the manner of the Indians hailing of a ship, calling it after the name of their *canoas*': babage canoaseen.

Arawak of 'near the mouths of the river Orinoco', V: jurace (iura . . . -ansi) *presto*, vacuna (bako-te) *ven acà*, cale (kali) *cazabe o pan*, vadahabò (bauhu, bahu) *casa*, cayure *camino*, iquigî (hikihi) *lumbre*, casacomuín (kasaku, cf. kasakabo *day*) *cielo*, aguadule (uaduli) *viento*, guine (oni- ~ uine) *agua*, curacale (akurakali, kulakani) *trueno*, abarùa (abaro) *1*, viama (biama ~ bian) *2*, cabuin (kabun) *3*, viauite (bibiti) *4*, aba dacabo (aba dakabo) *5*, viama dacabo (bian dakabo) *10*, soco *muger* (cf. sioko *small*), dadite (daditi ~ daiti) *mi hijo*, dajuna *mi padre* (cf. daionati *my parents*), dabuquei *mi amigo* (cf. dabukiti *my elder brother*), tapane *mi amigo*.

Nepuyo of Trinidad, V: amuere *a ti digo*, chare yerepare *dame de comer*, tamacare *aguarda*, ipura mana *no lo ai*, acayaremate *mientes*, quanane *digo verdad*, corpiase *yo*

me muero de hambre, yomeyomese *date priessa*, tamacaresea pone *yo me doi la que puedo*, guapot *lumbre*.

It should be noted that Trinidadian Arawak of 1595 and modern Lokono have the same descriptive terms for *scissors* and *spy-glass*, containing, respectively, stems meaning *cut* and *see* together with an instrumental suffix; such forms are unlikely to have been borrowed, or inherited by more than one Arawakan language. They also have the same word for *sun*, hadali, which seems to be without cognate equivalents in any other Arawakan language, and the same prefix dA- marking 1st person singular, cognate equivalents of which have been reported only for Lokono, Goajiro, Paraujano, and Taino.

De Laet's saeckee *mother* is most probably cognate with modern Lokono sake *uterus or egg-shell*, which is a compound of sa-*child or egg* with eke *container or clothes*. Similarly, Island Carib ákai (Dominican) and ágai (Central American) *container*, which are cognate with Lokono eke *container* (as are I.C. áika and áiga *eat* with Lokono eke *eat*), are sometimes employed with the sense of *mother*, or at any rate as the stem of verbs derived therefrom meaning *be* (*somebody's*) *mother*. So Breton's (XVIIth C. Dominica) "toka çaga nakayem *celle-cy est ma mere*" (nákaię *is my mother*); and from my own recording in British Honduras, ka bágaią? *who is thy mother?* It should be noted that in these languages the nearest equivalent to our *mother* also includes *mother's sister*.

DOMINICA, BRITISH WEST INDIES

NOTE ON ATHAPASKAN AND SINO-TIBETAN

ROBERT SHAFER

Five years ago the writer presented in this journal some parallels between Athapaskan and Sino-Tibetan.[1] At that time it was pointed out that circumstances were unfavorable for ever having many parallels between the two families.[2] The Athapaskan part of the comparisons below come primarily from the manuscript vocabularies from recordings of Edward Sapir or Harry Hoijer, which were kindly lent me by Hoijer. The numbers of the comparisons show where they fit into the phonetic pattern outlined in the previous paper.

1.12; 20.2. Chiricahua dò·dà' *no, not.* Sino-Tibetan *t'a *prohibitive.*

1.13; 20.5. Chir. tà· (mom.), tá (cont.), tą́, tà·ɬ, tà· (2) *look for, search* (tr. and intr.), but Kato -te *look for anything*, Mat. tχe·, -tχe', tχe·l *look for*. Old Bodish lta *to look.*

1.14; 20.10. Kato naL gi· *dog*, Mat. na·- ɬlí, Wail. na·-ti· Lolo: Lisu a-na, Lolopho a-no < *-na. For the last part of the Athapaskan compound cf. Digaro no-kwi; Rodong k'li.

1.15; 20.1. Mat. -ba-ye *under*. O. B. ma (root) *below.*

5.3? Mat. -o·-, -o- *3d pers. optative* (*or permissive*). Cf. O. B. -o- *imperative.*

7.3. Navaho 'é·', 'é· *coat, garment*, Mescalero 'é· *shirt, clothing*, Chir. 'é·' *coat, shirt*, Chip. 'í *coat, dress.* Chinese 'ęi *upper garment*; *clothes* (Karlgren 185).

7.4; 20.11. Nav. lè·' *in the earth, sand*, lè·j *sand, soil*. Kukish *ŋ-lei\, Lušei lei *earth, ground.*

10.6; 20.2. Nav. dèg *up, above* (Hoijer, N. Ph. 10). O. B. t'eg *lift, raise*, and cf. ądeg(s), pf. bteg(s) *lift, raise, elevate*, ltag *upper part, above.*

11.6; 20.2. Ath. *-gid, Hupa, Wail. -gid (2) *be afraid*, Mat. -dji' or -djid, -djid, -djil, -dji', Nav. -dzìd, -dzí·d, -dziɬ, Jic. -dzi', Chir., Lip., KA -dzì, Sar. -dji' (d-) (2), Chip. -djèr (2). O. B. (?) k'ri-le-ba *fear*, Kachin k'rit *to fear.*

[1] Athapaskan and Sino-Tibetan, IJAL 18.12–19 (1952).

[2] By a typesetter's error the "many" was left out, leaving the statement to read "one will probably never have parallels between the two families," whereas the writer's main purpose was to present some parallels.

12.3; 20.4. Mat. ɬa' *feather*. O. B. ądab-ma *wing*.

Alternation between a final occlusive and the corresponding final nasal does not occur frequently in Sino-Tibetan, but neither is it rare. We seem to have a case of this in the following comparison:

20.8. Nav., SC -zìd *liver*, Chir., Mesc., Lip. -zì, Jic., Sar. -zì', Hupa -sit', Chip. -ðə́r. O. B. mtš'in-pa,[3] Kanauri šin, Burmese -sàñ < *-sìn.

Ath. -s = ST -s in section 18 of the earlier paper. But Old Bodish had four initial sibilants (š, ž, s, z) but only one final sibilant (s) which disappeared in most languages. Sometimes O. B. -s is an inflection, and perhaps sometimes a suffix. In the first of the following comparisons the Athapaskan forms seem to be primitive and the Old Bodish a perfect form of the primitive verb (*la) used as a noun:

20.11. Chip. là *work*, Mesc. lá, là·, lí·ɬ, —, lá *make, create, do*. O. B. las *action, deed, work*.

20.4 Mesc. -k'os *neck*, Mat. -k'os, Chip. -k̄wo, Mont. é-kkɣoš, Hare é-kk̄o, Louch. oe-kko. Daic *ɣå'å, Siamese gå'å.

20.11. Cf. Mat. -le·, -le' *to sing*, Kato -le'. Tsairel lai *to sing*, and Bunan la-la *song*, L. hla *song*.

Hoijer's stem lists enable me to authenticate some of the Athapaskan comparisons in the previous article with additional data from other languages:

1.5. Mesc. -má *mother*, Chir. -má·'.

1.8. Mesc. gà·'-yè' *a crow*, Chir. gà·'-yè', gà·-dé.

1.10 Cf. also Chir., Mesc. dą́ *time past* (v. suf.).

3.11. Nav. ni *you* (F), Mesc. ndí·.

5.1. Mesc. -do' *be hot*. For the loss of ST -r-, cf. *be afraid*, the fifth entry in this article.

10.3. Mesc. -nì· *face*.

15.1. Mesc. -ką́' *be sweet*.

15.3. Mesc. kį', kìnè' *house*.

[3] t is not part of the root, but is due to the prefix.

16.2. Mesc. xé·ɬ *burden, pack*.

22.2 (p. 19). To Chip. -(ǹ)nív-é *nostril*, add Nav. -ní·', Chir., Mesc. -ní. From the comparison of Chinese ńi/ *ear* with O. B. rna-ba and similar forms in -a in other Tibeto-Burmic languages, I have suspected that -i in the Chinese word may be the remains of an old dual inflection. If so, it may account for the -i in the Athapaskan stem for *nostril* compared to the -a in Tibeto-Burmic for *nose*.

BERKELEY, CALIFORNIA

OCEANIA MONOGRAPHS

We have great pleasure in informing you that the first number of the OCEANIA LINGUISTIC MONOGRAPHS, the new linguistic series advertised in Oceania, vol. XXVI, No. 2, will be available soon.

The author of this number is Dr. A. Capell, Reader in Oceanic Linguistics in Sydney University, and it is entitled A NEW APPROACH TO AUSTRALIAN LINGUISTICS (4°, approximately 100 pages, wrappers, price 10/- Australian currency, postage 7d within Australia and the British Commonwealth, 1/3 for other countries).

This volume represents an attempt to treat the sum total of the languages of the Australian aborigines on a comparative basis, and a great part of the material presented is based on the author's own fieldwork amongst the aborigines. Information is given on a number of areas which have so far been entirely unknown linguistically, such as parts of Cape York Peninsula.

Orders for the issue are invited and should be addressed to: The Secretary, Oceania Linguistic Monographs, Department of Anthropology, University of Sydney, Sydney, N.S.W., Australia.

Numbers in preparation include, amongst others, a collection of articles on Philippine linguistics by several authors, a study of the Nivo (Nifilole) language, Reef Islands, Santa Cruz Archipelago, by Dr. S. Wurm, a phonetic study of the Wahgi Valley dialects, Western Highlands District, New

Guinea, by Fr. L. Luzbetak, a grammar of the Bandjaloŋ language, north coast of New South Wales, by Dr. W. E. Smythe, a grammar of the Njaŋumaḍa language, Western Australia, by G. O'Grady, a comparative study of languages of the Western District, Papua, New Guinea, by Dr. S. Wurm.

Studia Phonetica

A new journal will appear under the joint editorship of E. Blancquaert, Ghent, S. K. Chatterji, Calcutta, W. Doroszewski, Warsaw, E. Fischer-Jørgensen, Copenhagen, P. L. Garvin, Washington, A. Gemelli, Milan, H. Glinz, Zurich, L. Hegedus, Budapest, R. Jakobson, Cambridge, K. Jimbo, Tokio, M. Joos, Madison (Wisconsin), B. Malmberg, Lund, A. Martinet, Paris, R. Olesch, Cologne, K. L. Pike, Ann Arbor (Michigan), E. W. Selmer, Oslo, A. Sovijärvi, Helsinki, and F. Trojan, Vienna.

Studia phonetica will serve research and criticism in linguistic phonetics. Many important papers, today printed in not easily accessible publications, are now to appear in this Journal.

The Journal will appear annually in two double numbers of eight sheets each and will be published by Vandenhoeck and Ruprecht, Göttingen, Germany. The languages of the Journal will be English, French and German.

The Publishers will deliver 25 separata gratis to the contributor and offer a remuneration of 20.— DM per sheet. This remuneration will be changed to an adequate author's fee in the course of the following years.

Recent Publications

General Linguistics

Brown, Roger W. and Donald C. Hildum. Expectancy and the Perception of Syllables. Lg 32.411–419, 1956.

Giet, Franz. Kann man in einer Tonsprache flüstern? Lingua 5.372–381, 1955–56.

Herdan, G. Language as Choice and Chance. Groningen: P. Noordhoff, 1956. xvi + 350 pp.

McQuown, Norman A. A Linguistic Laboratory Serves Cultural Anthropology. AA 58.536–539, 1956.

Sjoberg, Andrée F. and Gideon, Morris Swadesh, and T. N. Sreekantaiya. Problems in Glottochronology. AA 58.296–308, 1956.

Sommerfelt, Alf. Language, Society, and Culture. Norsk Tidsskrift for Sprogvidenskap 17. 5–81, 1956.

Sommerfelt, Alf. Sémantique et Lexicographie. Norsk Tidsskrift for Sprogvidenskap 17.485–489, 1956.

American Indian Languages

Hoijer, Harry. Athapaskan Kinship Systems. AA 58.309–333, 1956.

Hymes, D. H. Na-Dené and Positional Analysis of Categories. AA 58.624–638, 1956.

Mickey, Barbara. Acoma Kinship Terms. AA 12.249–256, 1956.

Wallis, Ethel. Simulfixation in Aspect Markers of Mezquital Otomi. Lg 32.453–459, 1956.

Journals Received

America Indigena. v. 16, no. 4, 1956.
American Speech. v. 31, no. 3, 1956.
Language. v. 32, no. 3, 1956.
Lingua. v. 5, no. 4, 1956.

ıdum.
n of

rache
5–56.
ıance.
xvi +

Lab-
ology.

Morris
taiya.
AA

, and
Sprog-

exico-
Sprog-

Sys-

tional
4–638,

Terms.

Aspect
. Lg

6.
6.